JUMS

Volume 7, Issue 3, July 2022

JUNIOR MANAGEMENT SCIENCE

Published by Junior Management Science e.V.

Junior Management Science 7(3) (2022) 543-568

Junior Management Science

journal homepage: www.jums.academy

Pricing Sustainable Shipping of Coffee: Consumers' Preferences and Willingness to Pay for Emission Reductions and Offsets

Ulrich Bek

Technische Universität München

Abstract

The goal of this study is to explore whether decarbonization of maritime shipping and the full supply chain are valued in customer perception. Understanding consumers' preferences and willingness to pay (WTP) for sustainable maritime shipping of goods can provide opportunities to spread the costs of required sectorial changes. Decarbonization labels were compared to the Fairtrade and European organic label. A discrete choice experiment (DCE) with 299 participants was performed and supporting data was considered for an exhaustive description of preferences and WTP in the exemplary use case of filter coffee. The results indicate a significant WTP premium for all labels. On average, direct reductions of all supply chain emissions were valued at 2.82€ (all values per 500g of coffee) and are thus comparable in importance to the Fairtrade label estimated at 2.77€. Maritime shipping offsets, reductions and offsets for the full supply chain were valued at average premiums of 1.79€, 1.95€ and 1.89€. Organic labelling led to an average premium of 1.61€. A random parameter logit model with correlated parameters found significant preference heterogeneity across participants for all labels. Participants preferring whole bean to ground coffee did not significantly differ in their underlying preferences for the sustainability dimensions but in their price perception and effectively displayed higher WTP for all attributes. This study contributes to current research by providing a thorough measurement of preferences and WTP for emission reductions along the supply chain and is the first to assess offsets compared to direct reductions in a controlled setting for a common product use case.

Keywords: Sustainable maritime shipping; direct decarbonization; carbon offsets; discrete choice experiment; pricing.

1. Introduction

> "Everybody in the world benefits from shipping, yet few people realize it. We ship food, technology, medicines, and memories. As the world's population continues to grow, [...] efficient maritime transport has an essential role to play in growth and sustainable development. [...] Maritime transport is the backbone of global trade and the global economy."
>
> – Ban Ki-moon as United Nations Secretary-General (United Nations, 2016)

Maritime cargo shipping is a vital pillar of our globalized civilization. Relatively speaking, it is a comparably eco-friendly mode of transportation in terms of greenhouse gas (GHG) emissions per kilometer per ton of freight transported (Creutzig et al., 2014). In absolute terms however, the industry's mere size results in a substantial contribution to annual global emissions. In 2018 the sector's GHG emissions were estimated to add up to 1076 Mt carbon dioxide ($CO_{2)}$ equivalent, which corresponds to 2.89% of all annual global anthropogenic emissions (Faber & Kleijn, 2020). Mainly due to growing global trade, GHG emission increases of up to 50% are expected if no regulatory action is taken (Faber & Kleijn, 2020). But there is potential for a different scenario.

The industry is currently at a turning point. To reach the Paris Agreement (United Nations Treaty Collection, 2015) goals of limiting global warming to below 2 °C pressure on the relevant regulatory body, the International Maritime Organization (IMO), is increasing. In addition to institutional and regulatory pressures, market factors and social forces drive the IMO to act and lead the industry towards sustainable maritime transport (Serra & Fancello, 2020).

Psaraftis (2019) defines sustainable maritime transport as "striking the right balance between varied and potentially competing economic, social and environmental objectives".

DOI: https://doi.org/10.5282/jums/v7i3pp543-568

Maritime transport without causing any detrimental effects on the environment but lacking economic viability and consideration of social aspects does thus not qualify as fully sustainable.

This study primarily focuses on only one part of this definition. Only one sustainability criterion is in scope: GHG emissions. More specifically, it focuses on two categories of emissions, those caused by maritime shipping only, and those caused along the full product supply chain including maritime transport.

The GHG emission reduction efforts currently proposed by the IMO primarily revolve around three topics: technological measures (such as alternative fuels and renewable energy use), fleet-related operational and management measures (such as improvements in speed management and route planning), and market-based incentives that either "discourage the use of high-carbon fuels" or "encourage the adoption of low-carbon practices through incentives" (International Maritime Organization, 2020a, 2020b, 2020c; Serra & Fancello, 2020).

It will be challenging for the industry to comply with the expected regulatory changes estimated to cost the industry up to $60 billion USD per year (Pitt, 2017). Barriers to reach the IMO's emission goals include change aversion in the sector, investment-related risks, uncertainty about future regulations, information and time constraints, limits in the technological feasibility of measures proposed, as well as market-related issues and political obstacles (Serra & Fancello, 2020). In line with the beforementioned definition of sustainable shipping, the authors of a recommendable overview of these challenges conclude that, "the real challenge for the future is to succeed in effectively integrating environmental sustainability with economic sustainability and shipping needs" (Serra & Fancello, 2020).

An additional transitional approach to counteract the sector's environmental impact is to compensate for GHG emissions that cannot be prevented by funding measures to reduce emissions elsewhere (Meunier, Stoll, & Schoen, 2019). This approach called carbon offsetting is based on the idea that local emissions have global effects, and the sum of global emissions is what ultimately matters. It can thus be less effective to strive for zero emissions in one specific use case while neglecting other, more important emitters (Kollmuss, 2010). While the global market for voluntary and non-voluntary carbon offsets is growing rapidly, a lot of controversy is surrounding the approach for various reasons including ethical concerns and a lack of efficacy and efficiency of the offsetting measures taken (Day, 2021). The main critique on GHG offsets is summarized very well in a joint statement of multiple environmental organizations from 2006: "Purchasing offsets can be seen as an easy way out for governments, businesses and individuals to continue polluting without making changes to the way they do business or their behavior" (Canzi, Clough, & Kronick, 2006, p. 1). Direct reductions should thus usually be preferred to funding indirect reductions. Nonetheless, offsets can be a useful transition mechanism, but other policies can often result in faster and more efficient net reductions (Kollmuss, 2010).

This study contributes to aligning economic and sustainability objectives by addressing some topics that are currently (at least to some degree) neglected in literature. The three key issues in scope of this study are:

- understanding whether sustainable shipping is a relevant topic from a consumer perspective by estimating willingness to pay (WTP) for decarbonized maritime shipping and full decarbonization of supply chains,
- comparison of customer preferences for direct emission reductions with preferences for emission offsets, and
- comparison of the proposed (shipping) emission reduction labels and offset labels with established sustainability labels.

The general assumption that underlies this study is that in consumer perception goods shipped in a sustainable manner (causing less or no emissions) could be valued more than goods that were shipped conventionally. If this perceived value exists and influences decision behavior, this benefit can be quantified as a WTP. Estimating increases of WTP based on employing more environmentally friendly forms of shipping can enable innovations and business models built around this decarbonization and therefore act as a driver of change in the industry.

To test this assumption, customer preferences and WTP for compensation and reduction efforts are quantified in a discrete choice experiment (DCE) (Carson & Louviere, 2011) on a specific example use case: roasted ground and whole bean filter coffee. Coffee was chosen as the use case in scope for multiple reasons:

- Raw coffee is usually shipped by sea and then roasted near the target market. At least 80% of all German coffee demand is shipped by sea (Deutscher Kaffeeverband e.V. , 2020), as Europe lacks the climate to grow coffee[1].
- An estimated 61.8 % of the German population (14 years or older) consumes at least one serving of roasted filter coffee each day (Förster, 2020). Coffee is an omnipresent and frequently bought product. From a survey design perspective, a product like coffee that most participants use and should be acquainted with (in terms of common product features, prices, and packaging sizes) is preferable for accurate and valid results.
- Coffee suits the context of decarbonization well. As van Loo et al. (2015) point out, coffee can be considered a pioneering industry for sustainability certifica-

[1]The exception being one plantation in a suitable valley on the Canary Islands, which is geographically part of the African Plate but politically an autonomous region of the European Union.

tion schemes. It is a major industry for both environmental (e.g., organic) and social (e.g., Fairtrade) certification (Fairtrade International, 2021b). While carbon labels are still rather rare in the industry (van Loo et al., 2015) a niche of "Segel-Kaffee", sail-shipped coffee transported on sailing cargo ships does exist (Klein, 2021).

In addition to these theoretical advantages, a practical consideration has influenced the choice of coffee as a use case. This thesis was written in cooperation with the startup company CargoKite. With the goal of contributing to the decarbonization of the marine industry, CargoKite has developed a novel ship concept that makes it possible to use wind as the sole propulsion using a kite system. As this could allow them to achieve reliability and cruising speeds comparable to conventional cargo ships, their goal is to eventually compete with conventional container shipping at scale. This differentiates them from the few existing niche companies in the sector, for example EcoClipper (2021) that attempt to revive sailing cargo shipping on conventional, classical sailing ships in smaller scale niche markets.

An interesting beachhead market for initial CargoKite prototypes is (premium) coffee. With viable wind conditions on common coffee shipping routes and a stable global demand even in times of crisis (Deutscher Kaffeemarkt e.V, 2021; International Coffee Organization, 2021), (premium) coffee supply could be a viable first market to operate in, especially if consumers were willing to pay a moderate surcharge for products shipped sustainably to fund initial research and development cost. Thus, the practical relevance of this study lies in pricing the use case described to benefit CargoKite and other innovators driving decarbonization in comparable contexts.

An additional theoretical benefit of this study is that it provides first insights into whether emission offsets and direct reductions are valued differently by consumers. Furthermore, the study gives a quantified indication of the impact of creating a partially compared to a fully decarbonized supply chain, both for direct reductions and emission offsets. Influences of the Fairtrade (Fairtrade International, 2021b) and the European organic label (European Commission, 2021) on choice behavior are also measured in the study, mainly as a reference, to provide a realistic context for and indicate the plausibility of the above results. Of course, this research on the exemplary use case of coffee cannot provide a context-independent universally true value of WTP for sustainable (maritime) shipping, it does however provide a rigorously controlled exemplary estimate that can serve as a first indication for comparable settings of interest.

This study furthermore provides an applied example for performing a DCE with sequences of multinomial choices[2] (i.e., multiple discrete choices are made by each survey participant throughout different, statistically efficiently calculated choice sets and preferences are then derived from this data with a random parameter multinomial logit model), solely relying on[3] the free open-source statistics software R (R Core Team, 2021; RStudio Team, 2021). Both the generation of the choice design, as well as the analysis were performed using R and relevant packages (Croissant, 2020; Traets, Sanchez, & Vandebroek, 2020).

The study was performed online with a survey focused on the DCE. For context, individual specific variables (such as demographics, attitudes, comprehension measures and stated preferences) were also part of the survey.

In chapter two the presented work starts off with relevant theoretical background on the topics introduced above. A detailed record on the sample characteristics and experimental measures is provided in chapter three, followed by a thorough discussion of data analysis and the resulting findings in chapters four and five, respectively. After a brief discussion of limitations of this research in chapter six, the primary findings of this study are summarized in the conclusion (chapter seven).

2. Theoretical Background

The IMO has set ambitious goals to reduce their total annual GHG emissions by at least 50% compared to 2008 which requires reductions of approximately 85% per ship in operation (International Maritime Organization, 2020c). Reductions like these are only possible by broadly employing a mix of various measures. The IMO is currently in the process of defining and agreeing on short-, mid- and long-term measures, targeting 2023 to adopt and provide implementation schedules for the short-term measures that are deemed viable to reach their goals (Marine Environment Protection Committee [MEPC], 2018). This paper does not focus on the technical implications of implementing decarbonization in the industry. Nonetheless, to provide the necessary context to understand the issue at hand, the most important approaches to reduce emissions will be briefly summarized in the paragraph below. This summary is based on the work of Serra and Fancello (2020), a comprehensive overview of the measures as well as related challenges and opportunities.

As mentioned in the introduction, the main areas the industry focuses on to reduce shipping emissions are technological, operational (fleet-related), market-based and management measures. Most emissions are linked to fuel consumption. Today, the most common fuel is bunker oil (i.e., low quality diesel) causing relatively high emissions even in optimized modern marine engines. Thus, technological approaches revolve around alternative fuels with better emission performance, improvements in ship design efficiency, filtration of exhaust fumes and/or switching to alternative energy sources for propulsion like wind and solar energy. It

[2]This kind of study is sometimes referred to as a choice based conjoint study, as the term is ambiguous, this study will refer to methodology terms of Louviere's nomenclature instead (2011).

[3]Apart from the use of Qualtrics (2021), a service to host survey questions, all software used was open-source.

is currently uncertain which alternative will prove superior long term. Operational measures include better navigation and reductions in ship speed (also called slow steaming) and can strongly reduce fuel consumption and emissions. These measures can often be applied in short term but have the potential to conflict with other business goals e.g., by reducing overall transport capacity. Management and logistics-based approaches revolve around optimizing various specific processes related to maritime operations (such as berth allocation in ports) using simulations and mathematical modelling. Market-based measures are the third area of interest for decarbonization. They discourage the use of high emission technology and fuels (e.g., compulsory carbon taxes), or encourage the adoption of practices and technologies that result in lower emissions (e.g., subsidies for eco-friendly investments). Given the international complexity the shipping industry operates in, it is not considered likely to implement effective market-based mechanisms in the short but rather in the long term.

Serra and Fancello (2020) summarize multiple barriers towards implementing decarbonization measures at the required speed to reach the IMO's goals including:

- lacking technological maturity of some measures to function at scale;
- "chicken-and-egg" problems related to the changes required (e.g., ships using alternative fuels and the required infrastructure to provide these fuels lack operational viability without the other already in place);
- general risk aversion towards change and inertia of stakeholders;
- political obstacles (especially for market-based measures);
- economic barriers (such as required investment costs).

While many of these efficiency gains will also translate to better fuel economy, lower operational costs, and better economic performance (Raza, 2020), overall, compliance with the upcoming regulations the IMO proposes will be costly for the industry (Serra & Fancello, 2020). The question on how to share these investment costs across the ecosystems of those requiring shipping services remains unanswered (Egloff, Escudero, Sanders, Webster, & Zampelas, 2019).

The proposed study investigates an opportunity to share these costs by exploring whether consumers perceive value in decarbonized supply chains and are willing to pay a premium for sustainable maritime shipping. Comprehending sustainable shipping as customer value could be a solution or partial mitigator for the economic costs of change. This research focuses on quantifying the possible positive effects on WTP for products that were shipped in a sustainable manner in a DCE employing various forms of emission reduction and offset labels. Purely informational carbon footprint labels (e.g., showing the amount of carbon emissions caused without reducing or offsetting these emissions) are not within the scope of this study, while labels promoting offsets and reductions are. Thus e.g., the works of Akaichi, de Grauw, Darmon, and Revoredo-Giha (2016) and Steiner, Peschel, and Grebitus (2017) are considered out of scope due to carbon labelling with specific numbers (and their different use cases). Below, relevant customer preferences revealed in literature are explored resulting in the research question and hypotheses proposed.

2.1. Preferences and Willingness to Pay for (Shipping) Decarbonization

At the time of writing this thesis, to the best of the author's knowledge, no literature on WTP specifically for green maritime shipping practices on a product use case was available. Most research concerned with WTP for emission reductions and offsets is focused on other sectors such as residential energy (Streimikiene, Balezentis, Alisauskaite-Seskiene, Stankuniene, & Simanaviciene, 2019), aviation (Caputo, Nayga, & Scarpa, 2013; Schwirplies, Dütschke, Schleich, & Ziegler, 2019; Sonnenschein & Smedby, 2019), urban freight transport and road transportation (Lera-López, Faulin, & Sánchez, 2012; Polinori et al., 2018). Hence, selected, relevant results for WTP in different but possibly on some dimensions comparable settings are presented below. Delving into the various challenges of offset programs from a policy perspective exceeds the scope of this literature review, for an overview of offsetting programs, related risks, benefits, trends, qualification and certification issues please refer to Kollmuss (2010).

A study focusing on green transportation in general, not green maritime shipping was performed by Schniederjans and Starkey in 2014. Customer motivations, intentions and WTP to buy a t-shirt that was transported directly from "manufacturing [...] using a truck with energy efficient fuel" were explored applying the theory of planned behavior (Schniederjans & Starkey, 2014, p. 119). The study focuses on the motivational antecedents of WTP for green transportation. A frugal approach of directly asking for their participants' additional WTP for a "green transportation t-shirt" compared to a "general t-shirt" was employed, the results were analyzed descriptively based on multiple segmentation criteria and analyzed with structural equation analysis to model the impact of personal attitudes towards green transportation, perceived behavioral control and peer pressure on intention to purchase and WTP. Their results indicated significant effects of peer pressure and attitude on intention to purchase which in turn influenced WTP. There were significant gender differences, as males reported a lower additional mean WTP premium of \$4.06, 95% CI [\$3.58, \$4.55] than females, \$4.96, 95% CI [\$4.47, \$5.45][4]. Education, location, income, age and perceived behavioral control did not influence WTP significantly[5].

[4] Confidence intervals are reported in this overview if they were reported by the authors.

[5] Measured at $\alpha = 0,05$ significance level. A „significant" influence of age at $\alpha = 0,10$ is reported.

Polinori et al. (2018) have employed a similar approach on Italian university students in 2015, focusing on urban freight transportation and using a similar, vague "green transported t-shirt" label. The overall mean WTP premium for those who were willing to pay a premium was 4.86€ (SD = 3.3) per shirt. 155 out of 337 participants were not willing to pay a premium in an initial filter question. Females, as well as self-reported high-income as well as environmentally active participants and public transportation users were characterized by above average levels of WTP.

For more insights on the antecedents of individual WTP for (voluntary) carbon offsets, Tao, Duan, and Deng (2021) provide interesting insights based on modelling consumers' WTP for general voluntary carbon offsets using an extended theory of planned behavior. Specific knowledge about carbon offsetting and personal moral norms increased willingness to offset significantly by influencing attitude and perceived behavioral control. High-carbon consumers' willingness to offset was found to be strongly influenced by social status and social pressure concerns. Specific knowledge on carbon offsets and a detailed comprehension of the meaning of term carbon offsets cannot be assumed across the general population. Tao et al. (2021) reported low average knowledge of the concepts in their Chinese sample ($N = 905$). In the few previous studies published on the topic, average knowledge on carbon issues and carbon offsets in the US and Australia was also reported to be low (Polonsky, Garma, & Landreth Grau, 2011; Polonsky, Grau, & Garma, 2015).

Schwirplies, Dütschke, Schleich, and Ziegler (2017) examined the example of travel transportation, measuring the impact of individual factors and changes in framing on offset WTP. Participants with higher income, younger age and firmer environmental and politically social preferences were willing to pay more in the German sample ($N = 1005$). Individuals that believe in the efficacy of offsets for protecting the climate also show higher levels of WTP. Findings from Pleeging, van Exel, Burger, and Stavropoulos (2021) support this notion of the importance of efficacy beliefs, more specifically hope, as hopeful respondents from the Netherlands were more likely to pay more for emission reductions by sourcing green energy.

Besides individual factors, several external factors can improve the likelihood to voluntarily offset carbon emissions and/or increase the WTP for offsets. Blasch and Ohndorf (2015) differentiate between the WTP for offsets and the probability to offset in the first place. The first is predicted best by individual's internalized norms to avoid environmental degradation and partly on income. The latter is better explained by the expected social recognition for offsetting. Berger (2019) illustrate further framing effects on green products using signaling theory. Participants exhibited a higher WTP for green products when the product choice was public rather than private and the products signaled their sustainability clearly. Products that are costlier than their nongreen counterparts should thus be designed or labelled in a clearly recognizable way. Huber, Anderson, and Bernauer (2018, p. 235) reported that a combination of institutional signaling through publicly announced government policy (i.e., information about the Swiss government forcing industrial actors to offset) and group norm interventions e.g., "many of my friends are already voluntarily compensating their emissions" were most effective in persuading participants to express a willingness to offset and actual WTP. Similarly, matching of offsetting contributions by the travel provider, which could also be interpreted as a form of institutional signaling, also increased WTP in a different study Schwirplies et al. (2019).

Carbon label messaging that is framed as a gain-frame (e.g., "if you choose to offset your carbon emissions, you will be removing carbon from the atmosphere and helping to preserve our environment") results in significantly higher purchase intention of carbon offsetting products and significantly increased WTP, especially when combined with objective climate change information (e.g., "emission levels now exceed 400 parts per million, which has never occurred in the 800,000 years of recorded history") and objective carbon offsetting information (i.e., how the offsets are realized by whom) (Chi, Denton, & Gursoy, 2021, p. 5). Focus group discussions by Upham, Dendler, and Bleda (2011) resulted in the recommendation to use labels showcasing reduction efforts instead of specific emission values as they mean little to average consumers.

Regional projects and projects revolving around re-/afforestation were preferred to renewable energies and energy efficiency improvement projects by a German sample in 2019 (Schwirplies et al.). Ritchie, Kemperman, and Dolnicar (2021) reported a similar finding on the regionality of the projects for Australian air passengers but found that the freedom to choose a specific offsetting program does not increase WTP of air passengers. Baranzini, Borzykowski, and Carattini (2018) evaluated the acceptability of international in comparison to domestic reforestation offsets and found that emphasizing the cost-effectiveness of international reforestation programs can help reduce the preference for regional offsets. This preference might be contrary to rational economic incentives, as forestry projects are prone to risks that energy efficiency offsets do not inherit (e.g., reversal risks through natural disturbances such as insect outbreaks) (Galik & Jackson, 2009).

2.2. Preferences and Willingness to Pay for Sustainability Labels

Before diving deeper into the research question and experimental procedures undertaken, two important established labels also evaluated in this study will be briefly introduced: the Fairtrade label (Fairtrade International, 2021b) and the EG Bio Organic label (European Commission, 2021). Both labels are comparatively widespread for coffee and still grow in importance (Fairtrade International, 2021a; Willer, Trávníček, Meier, & Schlatter, 2021). An estimated 5% of all coffee sold in Germany in 2020 was certified as Fairtrade (Fairtrade International, 2021a) and approximately 4% of all coffee imported into the European Union in 2019 was certified as organic (Willer et al., 2021).

The main goal of the Fairtrade label is to "ensure fairer terms of trade between farmers and buyers, protect workers' rights, and provide the framework for producers to build thriving farms and organizations" (Fairtrade International, 2021b). Fairtrade is thus a predominantly economic and social label aimed at reducing international inequality and fighting poverty, its principles do however also incorporate certain environmental standards such as proper and safe management of chemicals, waste, soil and water resources. Fairtrade is widely recognized across industries and very present in the coffee industry. For example, 32,5% of all Fairtrade products sold were coffee products in 2019 (Fiedler, Frank, & Volland, 2020), thus it was chosen as an exemplary widely recognized label for social sustainability.

The second label that was part of the study is the European organic logo. It was introduced to give a coherent visual identity to European Union produced organic products. Several organic labels could have been chosen for this study, the EU organic logo was chosen because it is a mandatory label for all pre-packaged EU food products, produced and sold as organic within the EU (European Commission, 2021). It should thus be the most widely distributed and recognizable organic label for the sample. To give a detailed account of the detailed criteria for and criticism on both certifications exceeds the scope of this study. For reference please consult (Herrmann, 2015), for a recommendable overview over various German labels and their performance across multiple sustainability dimensions.

In contrast to the scarce literature on (maritime) transportation carbon reductions, multiple studies measuring the influence of established sustainability labels such as the Fairtrade and organic label (sometimes even of carbon neutral labels) on WTP in products use cases are available. The most relevant studies employing coffee as a use case and optimally estimating the impact of several labels in relation to each other will be briefly presented, to provide context for the results of this study.

Also, van Loo et al. (2015) have performed an experiment on coffee combining eye tracking with a discrete choice, evaluating the influence of visual attention on sustainability information. During their experiment WTP for coffee certified with USDA Organic, Fairtrade, Rainforest Alliance and a carbon footprint label was measured. The carbon footprint label, indicating "that the producer is reducing its carbon emissions", differs in content and is thus not objectively comparable to the carbon labels proposed for this study later. Based on a sample of 81 participants, the authors report the highest average WTP premium estimate of $1.16 per 12oz (∼340g) of coffee for the USDA Organic label, followed by the Rainforest Alliance and Fairtrade labels with $0.84 and $0.68, respectively. The carbon footprint label did not exert a significant influence on the choice behavior in their experiment.

A second study in a related domain was recently published by Birkenberg, Narjes, Weinmann, and Birner (2021) with results from a DCE performed in 2016 surveying a smaller sample of 80 German coffee bar visitors on their WTP for 250g packages of Arabica whole bean coffee. The packages varied on Fairtrade certification, a trust based direct trade without certification, as well as an offset based "carbon neutral" certification. The experiment resulted in a mean WTP premium of 1.77€ for carbon neutral offsetting, a mean WTP premium of 3.22€ for a non-certified direct trade claim and a mean WTP premium for Fairtrade certification of 4.30€ (all premia per 250g of coffee). The study furthermore contained insights from three focus group discussions on the topic. Interesting results included that while all groups mentioned fair trade relations as one of multiple important factors influencing their purchasing behavior, purely environmental aspects such as the carbon footprint had to be introduced by the moderator in every focus group and the participants showed little initial awareness of the environmental impact of coffee production. In an additional study focused on Fairtrade coffee Rotaris and Danielis (2011) reported an average WTP premium of 2.20€ for Fairtrade certification on a 250g package of coffee (beans or ground was not defined by the authors) based on a DCE of an Italian sample of 46 men and 89 women. Their own data and their review of previous studies in the domain indicated strong heterogeneity in premium prices across individuals. Factors explaining this heterogeneity included item type (e.g., ground coffee or brewed cups), age, gender, income, purchasing habits and individual views on ethical issues such as poverty and inequality.

Lastly, Lappeman, Orpwood, Russell, Zeller, and Jansson (2019) performed a DCE on Fairtrade coffee based on a South African sample ($N = 300$) incorporating personal values. They measured an overall mean WTP for Fairtrade of $1.22 per 250g of coffee (27% premium compared to the reference price). A cluster analysis based on the premium WTP was performed to analyze the influence of personal values. The only personal value showing a significant influence was humanitarianism i.e., the care for human welfare construct which was based on dimensions such as beliefs in equality and freedom. Besides this, few differences between the clusters were found, e.g., counter to the authors expectations, knowledge of Fairtrade did not differ between segments or influence WTP.

2.3. Research Question and Hypotheses

Considering this theoretical background, the research gaps this study addresses can be outlined. While there are multiple studies on WTP for established sustainability certification, labels promoting decarbonization are not yet as well understood. No study in the domain has been identified which evaluates consumers' perception of offsetting compared to direct reductions. Furthermore, this is the first study to evaluate the importance of sustainable maritime shipping to consumers. The main goal of this study is thus summarized in the following research question:

> **RQ:** What are consumers' preferences for price, emission offsets and reductions (for both maritime shipping only and the full supply chain)

as well as for Fairtrade and organic labelling in the context of 500g packages of filter coffee (ground or whole bean) and what is their resulting marginal WTP for these attributes?

This research question leads to the following testable hypotheses:

H1: On average, participants are willing to pay a premium for filter coffee that was shipped without causing net emissions, i.e., parameter estimates and the resulting marginal WTP premiums for direct emission reductions and offsets differ significantly from zero.

H2: On average, participants value decarbonization of the full supply chain more than decarbonization of parts of the supply chain (maritime shipping only), i.e., parameter estimates for the decarbonization of the full supply chain are significantly higher than the corresponding parameter estimates of maritime shipping only (for both, offsets and direct reductions).

H3: On average, participants attribute more value to direct emissions reductions compared to indirect offsets, i.e., parameter estimates for direct reductions are significantly higher than the corresponding parameter estimates for offsets (for both, maritime shipping only and decarbonization of the full supply chain).

H4: On average, participants that prefer ground or whole bean filter coffee have identical preferences for decarbonization, sustainability labels and price, i.e., relevant parameter estimates do not differ significantly between the participants prompted with whole bean and ground coffee.

2.4. Pricing Approach

An indirect choice based survey was chosen to test these hypothesis as it allows for a closer replication of realistic buying situations compared to direct estimation approaches (e.g., van Westendorp and contingent valuation) and is commonly viewed as the most suitable (but also one of the most complex and effortful approaches) to quantify customer preferences and WTP (Simon & Fassnacht, 2016). The process of identifying the most suitable category of pricing methods was guided by Simon and Fassnacht (2016). The DCE approach was chosen among different variants of choice-based approaches for multiple reasons. Firstly, it not only allows for relative preference statements between alternatives but also enables decisions to not purchase any goods, thus yielding more valid results (Simon & Fassnacht, 2016). Secondly, the method is recommended as it is characterized by reduced hypothetical and social desirability bias, especially for contexts including hypothetical and public goods relating to environmental topics (Drichoutis, Lusk, & Pappa, 2016; Norwood & Lusk, 2011).

Valuing the public good of mitigated CO_2 emissions as an attribute of a private product can further reduce hypothetical bias, as private products are less prone to this bias (Loomis, 2014). Contrary to popular belief and previous meta-analysis on the topic (Carson, Flores, Martin, & Wright, 1996; List & Gallet, 2001; Murphy, Allen, Stevens, & Weatherhead, 2005; Schmidt & Bijmolt, 2020) did not find evidence for the general, context-independent superiority of indirect valuation compared to direct approaches for private goods in terms measurement accuracy and hypothetical bias. The authors did however not differentiate between specific indirect and direct methods and modelling approaches, thus a properly executed and modelled DCE was still considered to be the most promising experimental estimation approach in the context. The recommendations by Hauber et al. (2016) and Hensher, Rose, and Greene (2015) informed all modelling decisions. The experimental approach was carefully chosen to potentially minimize hypothetical bias and is elaborated in more detail in the following chapters experimental procedures and data analysis.

3. Experimental Procedures

The following chapter will introduce the experimental measures taken throughout the experiment. The data collected for the main study consisted of the following measures presented in order of survey flow:

- preliminary in-store price comparison of ground and whole bean coffee
- stated buying habits of coffee products;
- socio-demographic characteristics;
- DCE;
- stated importance of relevant product aspects for coffee;
- an optional open question to address possible concerns about the CO_2 labels during the choice experiment;
- stated favorite coffee product with brand, product name, package size and average price (if such a product exists);
- a question battery to test the comprehension level of the difference between CO_2 compensation and direct reductions.

3.1. Preliminary In-Store Price Comparison

104 product offers from six stores in Munich, Germany, (two discounters, two supermarkets as well as one drug store and one organic store) were recorded from 24.02. to 31.03.2021. Product and brand names, grind (ground or whole bean), pricing, packaging size, Fairtrade and organic certifications were analyzed to inform the design of the choice experiment. The comparison showed that prices

varied widely between brands, roasts, and labels. The cheapest blends were sold at prices of 2.99€ to 3.49€ per 500g (e.g., "Ja! Kräftiger Röstkaffee") with the most expensive packages ranging up to approximately 15€ per 500g (e.g., "Martermühle Bio" product range). Whole bean coffee averaged at prices of 6.57€ per 500g while ground coffee averaged at 6.55€ per 500g, supporting the design choice to use the same price levels in both groups. The most common packaging size in the in-store comparison (500g) was chosen for the experiment.

3.2. Sample Characteristics and Recruitment

The German online survey was distributed from 12.04. to 29.04.2021. The resulting convenience sample was recruited through multiple channels including:

- the personal networks of CargoKite's founding team and the study's author,
- sharing on Facebook in various regional bulletin boards across Germany,
- untargeted sharing on LinkedIn and Xing.

Participating in the study was not incentivized. However, there is little evidence of positive or negative effects of incentives on response quality and sample composition in surveys (Singer & Ye, 2013). Of 462 total submissions, 436 contained data beyond the filter questions. A data check showed that some incomplete sets revealed signs of participants' fatigue (e.g., a single alternative was chosen throughout multiple choice sets). Submissions with complete choice data did not show any unfavorable patterns. As a precaution, submissions with incomplete choice data were excluded from the study. The resulting cleaned sample consisted of 299 submissions. This dropout rate of 31% might be considered higher than expectable for a survey of this length (Liu & Wronski, 2018) and will be discussed in the limitations section in more detail. The appropriateness of a sample's size depends on question format, choice task complexity, desired result accuracy, heterogeneity in the target population, respondent availability and whether subgroups separately analyzed (Bridges et al., 2011). Based on Johnson and Orme's (2003) sample size recommendations, the sample was appropriate in the context and given the DB-efficient design (discussed in detail the next section) a smaller sample size could have sufficed for valid results (Rose & Bliemer, 2013).

As shown in table 1, the sample was biased towards an above average education level and consisted of more females than males. Most of the sample stated to live in southern Germany. A filter excluded any participants that did not at least occasionally buy either ground or whole bean coffee. If both types were bought, participants were prompted with an additional question to confirm their most purchased type. Based on this data, they were assigned to one of two experimental groups, Ground Coffee or Whole Bean Coffee. Group demographics were mostly similar but relatively speaking group 1 (Ground Coffee) included more students and residents of Baden-Württemberg.

Both groups were presented with identical questions throughout the study, the exception being the labelling of the (otherwise identical) choice tasks either as ground or whole bean coffee to fit their preference. The inclusion of these two grind levels of coffee, two slightly different variants of one product, is in line with the study's goal to strike a balance between taking on a broad perspective on the research question, while also providing clearly specified, empirically grounded baseline information in the use case. Thus, the results can be useful as a reference for comparable contexts without compromising meaningfulness for the specific use case. Furthermore, by including both grinds, the recruitment of a bigger and more representative sample was possible, as ground roasted coffee has a 50% retail market share and whole beans accounts for an additional 37.3% of coffee sold in Germany in 2020 (Deutscher Kaffeemarkt e.V, 2021). Including two similar variants of the same product further gives a first indication of the robustness of the expected effects across slight changes in the context.

3.3. Design of the Discrete Choice Experiment

A DCE with multinomial choice questions and generic, non-labelled alternatives (except for the alternative specific no buy or status quo alternative) was performed.

The experiment was introduced using a direct translation of the cheap talk script proposed by Lusk (2003) to reduce potential bias in hypothetical valuation questions (cf. Appendix A). Clear descriptions of all attributes and levels are recommended to avoid comprehension issues and room for interpretation between subjects (Bridges et al., 2011). In the case of this study however, a balance had to be struck to give sufficient contextual information while avoiding unintended priming, social desirability and other confounding effects on participant behavior. The topic was introduced with concise information consisting of an explanation of the choice task and the cheap talk script to measure the decisions as neutral and as unbiased as possible. For example, emission levels were purposefully not explained to avoid emphasizing the differences between reductions and offsets, differences that participants might disregard in their natural choices. Inquiry into cognitive bias is not in scope of this study, for reference, Felser (2015) provides a straightforward overview of important psychological terms and biases mentioned in this study.

3.3.1. Presentation of the Experiment

The participants were presented with 14 predefined full profile choice sets in random order, each containing three alternatives including one no buy or status quo alternative. All profiles were introduced either as ground or whole bean filter coffee depending on group membership. The profiles were presented as simplistic 3D-rendered 500g coffee bags to mimic a realistic buying context (cf. Figure 1). To minimize unintended attention effects (cf. van Loo et al.'s (2015) study on the impact of visual attention on choice behavior),

Table 1: Socio-Demographic Sample Characteristics (in % of Segment, $N = 299$)

	Overall Sample	**Group 1: Ground Coffee**	**Group 2: Whole Bean Coffee**
Segment size	100.0	41.5	58.5
Gender			
Female	64.2	63.7	64.6
Male	35.1	36.3	34.3
Country of residence			
Germany	97.0	94.2	98.8
Austria	2.0	4.0	0.6
Other	1.0	1.6	0.6
County of residence			
Baden-Württemberg	54.8	42.7	63.4
Bavaria	29.4	37.9	23.4
Hesse	2.0	0.8	2.9
North Rhine-Westphalia	2.0	1.6	2.3
Other	8.0	12.9	4.5
Education level completed			
Secondary General School	2.7	1.6	3.4
Intermediate Secondary School	16.1	12.1	18.9
Grammar Schools (A-level)	25.8	24.2	26.9
Bachelor's degree	32.8	37.9	29.1
Master's degree	17.7	20.2	16.0
Other	3.3	1.6	4.6
Occupation			
Employees & civil servants	53.8	42.7	61.7
Student	32.1	42.7	24.6
Self-employed	4.3	2.4	5.7
Housewife or househusband	2.0	1.6	2.3
Retired	2.0	3.2	1.1
Other	3.0	3.2	2.9
Age (in years)			
Mean (*M*)	32.2	30.1	33.7
Standard Deviation (*SD*)	10.8	10.8	10.6

all labels were enlarged in same-sized boxes for better readability. A logo containing a globe with a transport route from America to Europe was added to emission attributes, to make sure all attributes were roughly equally visually appealing. Furthermore, all labels were colored in the same shade of green to ensure a similar contrast level. Figure 1 provides an original depiction of the German DCE as presented to participants, please refer to Appendix A for the original German survey questionnaire.

The order of the labels in the 3D-render was randomized at image creation. The order of the choice sets and the order of the profiles inside each choice set were randomized during the survey for each participant. The no buy (i.e., status quo) option remained on the right of the screen to reduce cognitive strain.

3.3.2. No Buy or Status Quo Alternative

A no buy alternative was added to create a more realistic setting (Rao, 2014), better mimic consumer choices and increase design efficiency (Brazell et al., 2006). The experiment was performed with a generic basic product excluding many value-adding elements (e.g., the package was not branded) and focuses on a new concept (carbon savings in shipping). Consequently, adding a no buy option was consid-

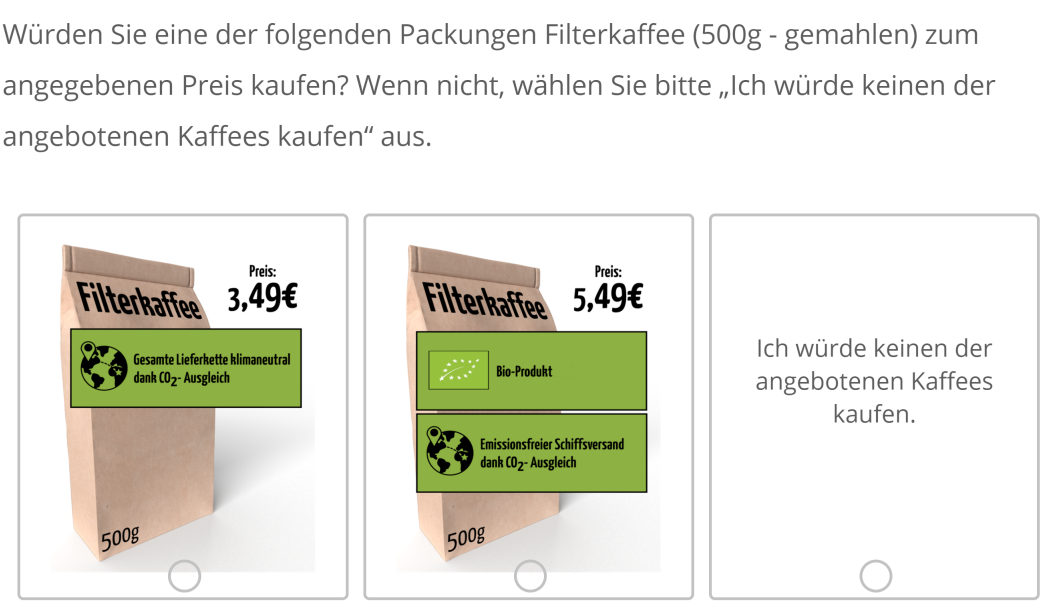


Figure 1: Depiction of a Survey Choice Set (Two Profiles & No Buy Option)

ered important to understand not only which alternatives and levels perform best in relative terms, but also whether they satisfy the participants minimum requirements for a (hypothetical) buying decision (Parker & Schrift, 2011). The disadvantages of adding a no buy alternative (e.g., lower information yield in sets where the no buy is chosen), were deemed less important than the above considerations, especially as most disadvantages can be mitigated by considering the no buy option correctly in the data analysis (Kamakura, Haaijer, & Wedel, 2001).

3.3.3. Attributes and Levels

The profiles differed on price (in € per 500g), emission labels, European organic certification, and Fairtrade labelling, as shown in table 2.

The continuous price levels were derived from the aforementioned in-store comparison and resemble common prices for ground and whole bean coffee. Both, the distance between price levels and the decimal places were kept constant between levels to avoid any confounding effects. Bliemer and Rose (2010) suggest that, while considering the risk of dominated alternatives, using a rather wide range is preferable to using a narrow range to avoid subjectively indistinguishable alternatives. As a result of these theoretical and practical considerations 4 steps of 2€ each from 3.49€ to 9.49€ were chosen as price levels. The grind was purposefully not added to the choice design as a level as it might not satisfy the basic DCE requirement of compensability (Gustafsson, Herrmann, & Huber, 2007). For example, a participant that does not own the equipment to grind coffee might never choose whole bean over bean, independent of the other attribute levels. The label level descriptions (organic, Fairtrade and carbon emissions) provide a simple, precise, and neutral account of the proposed value. Quantitative amounts of GHG emissions prevented/offset were not chosen as levels of the emission reduction attribute, assuming specific numbers mean little to average consumers (cf. Upham et al., 2011). Valuing the concepts of preventing and offsetting emissions is the goal of this study, while estimating WTP per ton of GHG emissions is not.

The categorical emissions attribute was defined in a balanced way, including four levels in total, two levels each for zero-emission maritime shipping and a climate neutral full supply chain, achieved by reductions and offsets, respectively. The number of levels for each attribute corresponds to the number of parameters that can be analyzed, therefore multiple levels were purposefully added as non-linear effects were expected for this attribute (Hess & Daly, 2014). The first carbon reduction level for zero-emission shipping was labelled with “Zero-emission maritime shipping thanks to wind energy” to describe the technology proposed by CargoKite as precise and neutral as possible. Although sail-shipped coffee is a somewhat more common term, it was not used to avoid evoking any emotional associations with sailing imagery. The second carbon reduction level was phrased “The full supply chain is climate-neutral thanks to emission avoidance with sustainable technologies”. Reducing emissions of the full supply chain to zero might not be technologically viable (yet), but the concept can already be described and the resulting WTP estimated. The corresponding offset levels were identically phrased, except for replacing the last part of the sentence with “thanks to CO_2-offsets”. The original German level descriptions can be found in Appendix B.

Only two specific labels (Fairtrade and organic) were added, to avoid inflating the design. Both labels are widely used on coffee and can be habitual buying criteria. Adding them provides a more realistic decision context. Furthermore, the labels’ results act as a reference to indicate the plausibility of the estimates in comparison with prior re-

Table 2: Choice Design with Profile Attributes and Levels

Level	Price	Organic	Fairtrade	Emission label
1	3.49€	No label	No label	No label
2	5.49€	Certified	Certified	Zero-emission maritime shipping thanks to wind energy.
3	7.49€	—	—	Zero-emission maritime shipping thanks to CO_2-offsets.
4	9.49€	—	—	Full supply chain is climate-neutral thanks to CO_2-offsets.
5	—	—	—	Full supply chain is climate-neutral thanks to emission avoidance with sustainable technologies.

Note: all prices in € per 500g of filter coffee

search. Adding the labels also allows for informative, comparative statements on the relevance and performance of the carbon labels relative to established labels for which market data and a richer body of existing research is available. The third reason for adding Fairtrade and organic labels was to provide an exemplary indication for the importance of social and production sustainability in contrast to GHG emission sustainability of maritime shipping.

3.3.4. Design Generation

The full list of possible profiles with these attributes and levels consists of 80 ($4 \times 2 \times 2 \times 5$) unique combinations. This full design was reduced to an efficient choice design using the modified Fedorov algorithm provided in the idefix package (Traets et al., 2020). The goal of this reduction was to achieve a set of tasks of feasible size that yields most information to estimate the parameters of interest. The algorithm optimizes the design for predefined parameter estimates while taking parameter uncertainty into account leading to a Bayesian d-efficient (so-called DB-efficient) design. Optimization is based on the DB-error i.e., the expected D-error minimized over the assumed prior distributions. This approach is in line with the recommendations of Traets et al. (2020) since designs optimizing for D-error are more sensitive to misspecifications of priors. Hensher et al. (2015) were referred to for theoretical guidance in defining the parameters and generating the design. The priors used for this optimization process were adopted from van Loo et al. (2015), who reported significant parameter coefficients for a price coefficient, organic and Fairtrade labels in a methodically and contextually comparable study. Using informative priors enables substantial improvements in the design's efficiency, and smaller ranges of the resulting parameter estimate confidence intervals in equally sized samples Carson and Louviere (2011). The parameter estimates for the emission labels were cautiously assumed to be zero. To maximize design robustness while taking participant fatigue into account, different set size specifications between nine and 16 choice sets per participant were estimated, following the common practice to include eight to 16 tasks per participant in a survey (Bridges et al., 2011). This iterative process resulted in a final design with a DB-Error of 2.56 consisting of 14 choice sets each containing two profiles and a no buy alternative.

3.4. Auxiliary Attitudinal and Knowledge Questions

To support the choice data, the stated importance of multiple possibly relevant attributes was evaluated using a 5-point Likert scale including a "no indication" answer option (cf. Appendix A). The DCE is focused on precise importance estimates of fewer product aspects, these ratings are intended to give a more exhaustive (although more frugal) overview over other influencing factors, their relative importance. The attributes and scales were inspired by Hasselbach and Roosen (2015) and modified by adding further buying criteria for coffee (e.g., brand, smell, origin, and taste) from a 2020 survey (Splendid Research) and attributes related to the research question like CO_2 impact. The order of the aspects was randomized between participants to avoid confounding positional effects.

To understand possible unintended influencing factors regarding the labelling or terminology used, an optional open question was added prompting participants to voice any concerns they might have towards emission-free or emission-compensation labels. This manipulation check monitors any negative attitudes towards emission reduction or compensation. For example, shipping emissions (although in absolute terms an important contributor to worldwide emissions Faber & Kleijn, 2020) might in relative terms be considered a small contributor in the lifecycle of coffee (Usva, Sinkko, Silvenius, Riipi, & Heusala, 2020). Psychological reactance could result from promoting it as a selling point. The measure was added after the choice experiment and the importance ratings to avoid unintended priming for concerns or consistency bias (Felser, 2015) in line with statements made.

An optional set of open questions was further added to evaluate stated product preference. Detailed information about participants' favorite (or habitually bought) coffee

product was requested. The product name and brand, followed by the packaging size and the average remembered price were only asked if they stated an existing preference. "No indication" options were included, packaging and price also contained an "unknown" option. These measures were intended to indicate habitual WTP and give a frugal indication of price interest and price knowledge.

As the last part of the survey, two quiz style batteries of variables were used to evaluate participants' comprehension of the difference between indirect emission compensations and direct emission reductions. Descriptions of situations were proposed, and participants had to choose whether these described CO_2 compensation or direct emission reductions, respectively. In total, six statements were presented for both tasks, two wrong statements and two statements that matched one of both definitions respectively. The order of the question batteries and the order of the items inside each battery were randomized for each participant to avoid confounding positional or learning effects. Prompting the participants to instead define the concepts and evaluate the answers manually was considered but refrained from to avoid straining participants and ensure objectivity of the results.

3.5. Data Analysis

There are multiple methods to conduct the statistical analysis of DCE data. This chapter will first introduce the software used to perform the analysis underlying this study. Secondly, it will give an overview over the choice model chosen and the assumptions that underly this model. Lastly, further complementary descriptive analysis undertaken in the context is introduced.

3.5.1. Software

The analysis was performed using R 4.10 "Camp Pontanezen" (R Core Team, 2021) in RStudio 1.4.1106 (RStudio Team, 2021). The packages tidyr (Wickham, 2021) and dplyr (Wickham, Romain, Henry, & Müller, 2021a) were used for general data transformation, epiDisplay (Chongsuvivatwong, 2018), ggplot2 (Wickham, Romain, Henry, & Müller, 2021b), texreg (Leifeld, 2013), Hmisc (Harrell, 2021) and skimr (Waring et al., 2021) were used for data visualization and summarization, idefix (Traets et al., 2020) and dfidx (Croissant, 2021) were used for generation of the choice design and formatting of the choice data and mlogit (Croissant, 2020) was used to perform the analysis and estimate all models. Finally, the rmarkdown (Allaire et al., 2021; Xie, Allaire, & Grolemund, 2018; Xie, Dervieux, & Riederer, 2020) package was used to create a reproducible workflows of all analysis performed. While advanced choice modelling is usually performed using paid specialized software, this work showcases that proper analysis adhering to scientific standards is feasible solely relying on free, easily accessible open-source software.

3.5.2. Analysis of the Choice Experiment with a Random Parameter Model

Multiple variations of models ranging from simpler multinomial fixed effects models to mixed logit models with various underlying assumptions were calculated for this study. The best and most suitable model to describe the data was a random parameter logit model (also called mixed logit model) taking the samples panel data structure into account and assuming all parameters to be normally distributed and correlated.

The random parameter logit model assumes that parameters vary from one individual to another and takes possible heterogeneity in the population into account (Croissant, 2020; Hauber et al., 2016; Hensher et al., 2015). As 14 repeated observations in the different choice sets were made, this longitudinal information was considered by assuming each individual can be described with a constant random parameter across all choice situations (Croissant, 2020; Hensher et al., 2015). The model considers the random parameters of individuals as random draws from a distribution whose parameters are estimated (Croissant, 2020; Hensher et al., 2015). It is necessary to define the properties of these random draws prior to performing the analysis. Pseudo-random draws based on Halton sequences were chosen, as these intelligent draws outperform truly random draws for simulation purposes in terms of stability and computational efficiency (Hensher et al., 2015). Multiple models with different ranges of draws from 100 to 2000 draws were estimated to confirm the stability of the final model (Hensher et al., 2015) and while not all models converged to completely stable solutions at first, neither the key results nor the parameter estimates changed drastically in later models. This lengthy estimation process is a commonly encountered problem for researchers employing the random parameter logit model (Hauber et al., 2016). The final model was created using 1000 (Halton) draws, as this (slightly) outperformed both models with more and less draws. Due to computational limitations only models with up to 2000 draws could be computed. In total, 20 models were computed ranging from 50 to 2000 draws and the best model based on the Log Likelihood value was chosen. The Akaike information criterion would have been an alternative measure of the goodness-of-fit of the model, correcting for complexity based on the number of parameters in the model (Field, 2013). As the same number of parameters were estimated in all models, the Log Likelihood value has similar informational value and was thus used to inform the model choice.

A Wald Chi-Squared Test from the mlogit package (Yves Croissant) was performed and concluded in a significant result, χ^2 (21, N = 299) = 653.11, p < .001. It tests the null hypothesis that the random effects are uncorrelated, the significant result thus indicates that the random parameters are correlated, supporting the choice of assuming correlations among the parameters. The resulting final model significantly outperformed all models assuming uncorrelated parameters in a Likelihood-ratio test comparing the correlated

model to the best uncorrelated model χ^2 (-21, N = 299) = 178.12, p < .001. The most suitable random parameter distribution for models with correlated parameters, a normal distribution, was chosen in modelling (Croissant, 2020).

To avoid confounding effects based on unobserved preference heterogeneity, all parameters were added to the model as random parameters (Hensher et al., 2015). In the resulting model each parameter is described with a mean[6], the estimated confidence intervals of this mean and a separate standard deviation of the distribution of the random parameters. This standard deviation, if significant, gives an indication of significant heterogeneity of the estimates in the sample. If heterogeneity is present in fixed effect models, this would lead to a reduced model fit but could in theory be handled through data segmentation (Hensher et al., 2015). For this study, an alternative approach could have been to create different fixed effect models for sample segments that exert similar choice behavior. In practice however, multiple reasons made the proposed model better suited in the context. Firstly, it would be unfeasible to measure and pick all the right segmentation criteria to explain this preference heterogeneity in a single study of this scale. Secondly, the given sample could become too small for meaningful and accurate results if divided into multiple segments. Thirdly, in the context of the study's scope a more realistic, general estimate is more meaningful than multiple segment-based estimates.

Another general benefit of mixed logit models is that they are not affected by the so-called independence of irrelevant alternatives property and thus provide a more complex but also more realistic model that does not assume that "the ratio of the choice probabilities is independent of the presence or absence of any other alternative in a choice set" (Hensher et al., 2015, p. 479). There was therefore no need to perform a Hausman-McFadden test for independence of irrelevant alternatives in this study (Hensher et al., 2015).

For a detailed mathematical derivation and description of the mathematical unconditional probabilities function that underlie the analysis performed, please refer to Croissant's (Croissant, 2020) for a concise overview, or to Train (2009) the original author of the theoretical content Croissant (2020) has applied in his work. The formulas used in the analysis are explained in much detail in the paper's chapter „5.1 Derivation of the model" more specifically in the section "Panel data", thus they were not added here to avoid redundancy.

The final model estimates the following parameters:

- The price attribute coded as a continuous variable.
- All other attributes, coded as dummy variables:
  - organic label;
  - Fairtrade label;
  - reduction of shipping emissions label;
  - reduction of full supply chain emissions label;
  - offset of shipping emissions label;
  - offset of full supply chain emission label.
- The no buy or status quo option was added to the analysis as an additional dummy variable to avoid a possibly lower model and predictive fit for the other (linear) attributes (cf. Kamakura et al. (2001) for more information).
- Lastly, interaction effects between group membership of group 1 (whole bean coffee) and all other parameters were added. These estimates reveal differences between this subsegment of the sample and the rest of the sample (if applicable).
- The intercept was omitted as the alternatives were not labelled (except for the No buy alternative that had a separate parameter) and an estimate of alternative-specific effects was thus not required.

The resulting WTP and WTP confidence intervals for the various labels in the WTP space was estimated based on the random parameter's marginal utility for each participant, dividing it by the corresponding individuals price coefficient estimate. Multiple WTP mean estimation techniques, e.g., the delta or Krinsky Robb methods as proposed by Hole (2007), were also considered, but given the availability of all required individual specific estimates the resulting WTP distribution could be directly calculated and the confidence intervals and properties of the resulting distribution were reported.

Several of the initial hypothesis can be answered by testing for equality of the parameter coefficients resulting from the model discussed above. The appropriate Z-tests were used to test for significant differences between these model coefficients (Paternoster, Brame, Mazerolle, & Piquero, 1998).

3.5.3. Further Analysis to Support the Discrete Choice Experiment

Further variables were analyzed to provide context for the results of the study, the following demographic data was analyzed descriptively to characterize the sample overall and divided by the two groups whole bean and ground coffee:

- age;
- gender;
- country of residence;
- county of residence in Germany (if applicable);
- education level;
- occupation;
- price and preferred package size of their favorite coffee.

[6] This mean is identical with the median, as the random parameter distribution is symmetric.

Furthermore, a manipulation check to monitor possible concerns that could cause psychological reactance towards the labels (Felser, 2015) was introduced. The qualitative data resulting from this optional open question was analyzed in a structured manner, guided by the Grounded Theory approach (Corbin & Strauss, 2015), to derive bottom-up descriptive categorical codes based on the data.

Comprehension of differences between offsets and reductions was analyzed based on the comprehension question battery data, descriptively, overall and by group, for offsets and reductions respectively. A score was calculated by adding a point for each statement matched correctly with the definition in each task. If a statement was matched falsely in one of the tasks, all points were lost for this task. This results in a scale from 0 (either no statement was correctly matched, or a mistake was made) to 2 (all statements were correctly matched without a mismatch). This mode of scoring was chosen as it allows for meaningful interpretations of the mean, which equals to the average number of tasks completed correctly. Differences between scores of groups of individuals were analyzed using non-parametric tests.

The stated Likert scale importance ratings for coffee attributes were treated as ordinal variables and analyzed descriptively using mean ranks, rank standard deviations and histograms. Correlations between the importance ratings, age, sex, and group membership were calculated using a Spearman's Rho correlation matrix to identify associated factors and reveal possible group differences in more detail.

4. Results

In the following chapter the study's results are presented starting with the analysis of stated importance ratings, followed by brief analysis of the in-store price comparison and learnings from habitually bought coffee in the sample. Conclusions from the manipulation check for greenwashing and the results of the comprehension check for offsets and reductions are presented. In this supporting context, the results of the DCE are introduced. Finally, in the following chapters five and six, these results are reviewed and discussed in light of existing literature and challenged in consideration of the study's limitations.

4.1. Stated Importance of Attributes and Consumption Preferences

The self-stated importance ratings of coffee attributes are summarized in Table 3, frequency bar charts for these attributes used for visual inspection can be found in Appendix C.

The main takeaways from this data were that taste seems to be the most important criterion for coffee overall, it was on average ranked much higher than the other criteria as "important" to "very important". Brands were rated as least important by the sample. The product attributes in scope of this study were ranked at medium importance. In general, the validity of this rank order should not be overestimated as the mean rank standard deviations are rather high indicating heterogeneity of preference in the sample. Plotting the sustainability ratings, all label histograms have a positive skew (most individuals rated them as important), with individual rating ranging from irrelevant to very important (cf. appendices C5, C7 and C8).

4.1.1. Exploratory Correlational Analysis of Stated Importance Ratings

The ranking data was analyzed exploratively in a correlational analysis employing a Spearman's Rho correlation matrix from the package Hmisc (Harrell, 2021). Insights from this analysis were very coherent with expectations and revealed multiple small differences in preference between the groups for whole bean and ground coffee. Dimensions relating directly to sustainability of coffee were correlated with medium effect sizes. For example, stated importance of the CO_2 footprint correlated with the stated importance of the organic label, $r_s = .41$, $p<.001$, $n = 293$. It also correlated with the rating of Fairtrade certification, $r_s = .43$, $p<.001$, $n = 292$, and the importance of product origin, $r_s = .33$, $p<.001$, $n = 288$. Interest in any sustainability criterion thus seems to be associated with interest in other sustainability criteria. When analyzing group membership effects, small differences are revealed. Group membership did however not correlate significantly with the importance of any sustainability dimensions[7] or with the reported gender.

Group 2 (whole bean coffee) did however slightly differ from group 1 (ground coffee) in age, importance of price and dimensions directly related to coffee quality:

- Group 2 (whole bean) participants were (significantly) older (weak effect) than participants in group 1: $r_s = .21$, $p < .01$, $n = 286$. As a reference, the mean age in group 1 was $M_1 = 30.1$ compared to $M_2 = 33.7$ in group 2 with standard deviations of $SD_1 = 10.8$ and $SD_2 = 10.6$ respectively.
- The second group's stated importance of price was (significantly) lower (weak effect) compared to group 1: $r_s = -.14$, $p = .013$, $n = 294$)
- Smell was rated as (significantly) more important (weak effect) in group 2 compared to group 1: $r_s = .23$, $p < .001$, $n = 291$.
- Taste was rated as (significantly) more important (weak effect) in group 2 compared to group 1: $r_s = .17$, $p<.001$, $n = 289$.

[7] Importance of Fairtrade label not correlated with group 2 membership: $r_s = -.02$, $p = .73$, $n = 293$ Importance of CO_2 footprint not correlated with group 2 membership: $r_s = -.02$, $p = .71$, $n = 291$ Importance of organic label not correlated with group 2 membership: $r_s = -.02$, $p = .79$, $n = 292$ Importance of product origin not correlated with group 2 membership: $r_s =$, 11, $p = .07$, $n = 288$ Furthermore, no correlation of gender with group 2 membership: $r_s = .02$, $p = .77$, $n = 297$

Table 3: Importance Ratings for Coffee Attributes in the Sample

	Rank order	Mean rank (SD)	*n*
Importance of:			
Taste	1	4.65 (0.74)	289
Digestibility	2	4.01 (1.16)	292
Previous product experiences	3	3.97 (1.07)	292
Smell	4	3.96 (1.07)	291
Fairtrade certification	5	3.62 (1.10)	293
Price	6	3.45 (0.98)	294
Carbon footprint	7	3.41 (1.09)	291
Organic certification	8	3.35 (1.10)	292
Type of roast	9	3.34 (1.24)	289
Origin	10	3.15 (1.15)	288
Convenience	11	2.83 (1.27)	282
Packaging	12	2.71 (1.12)	294
Brand	13	2.44 (1.17)	294

- The type of roast was also rated as (significantly) more important (weak effect) in group 2 compared to group 1: $r_s = .22$, $p<.0001$, $n = 289$.
- Coherent with expectations, as an additional step (grind) is needed to brew whole bean coffee, group 2 accredited (significantly) less importance to convenience (small effect) than group 1: $r_s = -.16$, $p<.001$, $n = 282$.
- Lastly, digestibility was also rated as slightly (but significantly) more important in group 2 than in group 1: $r_s = .15$, $p = .013$, $n = 292$.

4.1.2. Stated Habitual Product Preferences

45% of the 299 participants stated a favorite or habitually bought coffee product, with 136 participants indicating that they have no favorite coffee and 33 participants skipping the question. 69% of these 136 participants gave both a packaging size and a price estimate giving frugal indications of price interest and knowledge. Thus, the following data (see Table 4) describes only a subset of the sample but could give an indication of possible differences between the groups, nonetheless.

Welch Two Sample t-tests were performed to test if the two groups differ significantly on their mean price per kg and mean package size in g. While the reported price per kg did not differ significantly between the groups $t(81.57) = 1.60$, $p = .11$, the average package size did, $t(108.35) = 6.80$, $p<.001$, indicating that the whole bean coffee buyers were used to significantly larger packaging sizes than the ground bean buyers. As the chosen packaging size of 500g for the choice experiment represents a compromise between size expectations for both groups (and lies within a standard deviation for both), it proved to be a reasonable and realistic choice for the experiment. Qualitatively, few brands were cited multiple times (e.g., Lavazza, but also Darboven, Hochland, Mövenpick and Tchibo) and a varied mixture of local roasters, discounter home brands, online shops, and organic specialty coffees was mentioned hinting at diverse preferences. As a conclusion, stated preferences about favorite coffee products gave first indications of existing diversity of brand preference, price knowledge and interest in the sample.

4.2. Results from the In-Store Price Comparison

The sample of coffee packages was too small for meaningful differentiation of the effects of organic and Fairtrade certification in consideration of interaction effects as the coffees were often accredited with multiple labels. Controlling for confounding variables such as brand and roast would not have been possible given the sample's limited size. As a first indication of the magnitude of expected effects, on average the 35 certified products with at least one label were sold at higher prices ($M = 8.32$€ , $SD = 3.32$€) than the 70 products without organic or Fairtrade certification ($M = 5.69$€ , $SD = 2.75$€). A Two Sample Welch-test was computed to test whether this difference is statistically significant while accounting for the unequal sample sizes of the two groups and found a significant difference between the prices of certified and uncertified products, $t(56) = 4.02$, $p<.001$.

4.3. Qualitative Analysis of Reported Concerns

As previously discussed, a manipulation check for unintended reactance towards the emission-free shipping labels was added. 60 individuals voiced concerns, 30 were members of groups 1 and 2 respectively. Some submissions contained information relating to multiple analysis categories. Two main categories of concerns were voiced: insufficient description and the credibility of the statements. Some concerns were mentioned that did not relate directly to the issues at hand (e.g., "wind power plants also impact our ecosystem"[8]) or were hard to allocate due to insufficient informa-

[8]Exemplary original comments were translated trying to minimize changes in style or meaning.

Table 4: Description of Favorite* Coffee Products by Segment (n = 107)

	Overall Sample	**Group 1: Ground Bean**	**Group 2: Whole Bean**
Segment size in %	100	36	64
Package size in g:			
Mean	688	442	830
Standard Deviation	389	202	401
Price per kg in € :			
Mean	18.1	16.2	19.1
Standard Deviation	9.3	8.73	9.44

* The sample indicated their favourite or habitually bought coffee.

tion provided by the participants. The data was structured inductively, and not fitted onto an existing model resulting in the partially ordered overview of concerns provided in Table 5.

The first category "insufficient descriptions" can be further specified with three issues that emerged multiple times:

- lacking general comprehension of the meaning of the labels (e.g., "honestly, [I] cannot imagine what is meant and thus not assess the importance");
- lacking transparency on specific aspects of the labels and wishes for clarification of distinct aspects (e.g.,"[the label] suggests transport on a sailing ship (if that is the case: concrete description + certification! This way higher prices would be by all means acceptable");
- missing clarity of the logical link between powered by wind energy and shipping was also expressed by some participants (e.g., "[I] don't understand how something can be shipped by wind power [..]"), although this is also a specific transparency issue, as it was voiced by various participants it is presented as an additional subcategory. The concept of sailing might not be known to all participants, counter to the researcher's assumption.

The credibility concerns are specified further through multiple subcategories:

- concerns with the general credibility of certification and distrust towards labelling schemes (e.g., "[I] have little faith left in certifications and labels");
- greenwashing concerns (e.g., "If it is called emission-free but then restricted, this smells like cheating/ greenwashing"), concerns like these are known to cause consumer behavior that is opposed to the goals of affected companies or organizations (de Jong, Harkink, & Barth, 2018);
- uncertainty about the relevance of shipping compared to the total carbon footprint of coffee in the context of the product lifecycle (e.g., "I don't know how big the total effect is: E.g., shipped emission-free thanks to wind power might equal a 50% reduction of carbon in the whole value chain?");
- general disapproval of carbon offsetting practices (e.g., "I feel that offsetting emissions with other sustainability projects is sometimes an excuse not to deal with them in the product lifecycle";
- doubts on the technical feasibility of shipping without emissions at our current level of technology or given the current state of the industry (e.g., "Global transport of goods is not yet possible (completely) emission-free").

The main goal of this manipulation check was to assess whether a majority of participants showed reactance to the labels presented because of greenwashing concerns or failed to understand the tasks. However only 2% of the sample did report greenwashing concerns and only 2% of the sample reported confusion about the labels. Participants voicing these more serious concerns were not excluded to reflect the sample's preference diversity. Further questions that arose because of the purposefully frugal description of the context were expected. With 7% of the sample voicing slight needs for more clarification, this is certainly a valid point to consider in further research but given the general focus of this research the labels seem to have been described comprehensively.

4.4. Comprehension of Reduction and Offsets

Comprehension of the differences between reductions and offsets was tested with two quiz-style batteries. 94% of the sample took part in this quiz located as the last part of the survey. Of these 281 participants, only 16% were able to identify all statements and assign them to direct reductions or carbon offsets correctly.

Two variables were calculated to describe the comprehension of offsets and the comprehension of reductions, respectively. The sample scored higher when assigning examples to the direct reductions ($M = 1.14$, $SD = 0.83$)

Table 5: Frequency of Concerns Voiced in the Sample ($N = 299$)

Description of concerns on emission-free shipping	Frequency count	Percentage of sample
Total count of concerns voiced	60	20%
Insufficient description of the context:	22	7%
Lacking transparency on specific label aspects	8	3%
Lacking general comprehension of label meaning	7	2%
Unclear link between shipping and wind power	7	2%
Credibility concerns:	32	11%
General credibility of labels	11	4%
Greenwashing* i.e., marketing vs. altruistic motives	6	2%
Relevance of shipping vs. total carbon footprint	6	2%
General disapproval of carbon offsets	4	1%
Technical feasibility of shipping without emissions	4	1%
Unclear and unrelated concerns:	10	3%
Unclear information provided by participants	5	2%
General feedback without mentioning concerns	3	1%
Attitudes on non-related issues	2	1%

*Greenwashing describes making „people think that [an organization] is concerned about the environment, even if its real business [...] harms the environment" (Oxford University Press, 2021))

than to carbon offsets ($M = 0.65$, $SD = 0.8$). As the two variables did not visually resemble a normal distribution, a paired Wilcoxon signed-ranks test was performed to test if scores of both tasks differed significantly in their central tendency. Indeed, the participants' comprehension score for direct reductions was significantly higher compared to the comprehension score of carbon offsets $T = 9386$, $p<.001$. Two independent Wilcoxon rank-sum tests showed no significant differences in performance between groups one and two in terms of comprehensions of reductions ($W = 9309$, $p = 0.58$) and offset comprehension ($W = 10042$, $p = 0.53$). These results indicate that most participants did not have a perfect understanding of the definitions of the terms and that offsets were harder to understand than direct reductions.

4.5. Modelling Consumer Preferences using Random Parameter Logit

Table 6 shows an overview of the results of the final model summarized on the next page. In total, eight parameters and eight corresponding interaction effects were estimated. The random parameter logit model estimates multiple metrics for each parameter. The mean values of the parameters (β_p) describe the average part worth estimate across the population of respondents (Chapman & Feit, 2019). The standard error of the mean indicates the average deviation of this parameter estimate from the true population average value and thus measures the precision of this estimate. The next column shows the standard deviation of this mean (SD_p), signaling how the parameters varied across the sample's population between individuals (Chapman & Feit, 2019). The corresponding standard error of this estimate is provided. The standard deviation is a measure for the existence of homogeneity or heterogeneity of preferences between individuals belonging to the sample's population.

The significance levels of all estimates are shown by the asterisks (cf. legend below the table). The Significance of the mean estimates (β_P) indicates whether they deviate statistically significantly from zero. Significant estimates thus exerted a significant impact on the choices made. Significance of the standard deviation (SD_P) indicates that the coefficients vary significantly in the population and that a mixed logit model provides a significantly better representation of the given choice context than a comparable multinominal fixed effects model assuming no preference heterogeneity could. The size of all estimates needs to be interpreted in relation to the other parameter estimates in the model. To enable a more applicable interpretation in absolute terms (and comparability to other contexts), WTP values based on the parameter in scope and the price coefficient estimate were calculated. The resulting WTP values are presented, visualized and discussed in chapter 4.5.1.

The model shows that all parameter estimates have a statistically significant influence on the choices made by the participants. All labels had a positive impact on choice. If they were present, all other things equal, participants were more likely to choose the given alternative.

In relative terms, Fairtrade certification ($M = 3.69$, $SE = 0.09$) and the reduction of all supply chain emissions ($M =$

Table 6: Random Parameter Logit Model with Coefficient Estimates ($N = 299$)

Parameter	**Mean (SE)**		**Standard Deviation (SE)**	
Fairtrade label	3.69***	(0.09)	2.75***	(0.18)
Organic label	2.20***	(0.17)	1.86***	(0.16)
Offset maritime emissions	2.50***	(0.21)	1.84***	(0.21)
Offset full supply chain	2.73***	(0.30)	2.14***	(0.26)
Reduction maritime emissions	2.61***	(0.31)	1.97***	(0.25)
Reduction full supply chain	3.67***	(0.35)	2.06***	(0.28)
Price coefficient	-1.33***	(0.09)	1.02***	(0.07)
No buy coefficient	-3.08***	(0.46)	6.85***	(0.42)
Interactions terms with Group 2: Whole Bean				
Fairtrade label x G2	0.13	(0.21)		
Organic label x G2	-0.12	(0.09)		
Offset maritime emissions x G2	-0.44	(0.29)		
Offset of full supply chain x G2	-0.23	(0.33)		
Reduction maritime emissions x G2	-0.39	(0.34)		
Reduction full supply chain x G2	-0.55	(0.36)		
Price coefficient x G2	0.62***	(0.09)		
No buy coefficient x G2	2.00***	(0.49)		
Akaike's Information Criterion				4882.04
Log Likelihood				-2389
Num. obs.				4186
K				3

Note: $***p < .001$; $**p < .01$; $*p < .05$

3.67, $SE = 0.35$) were similar in strength and were the most important predictors of choice behavior. The coefficients for decarbonization of maritime shipping with reductions ($M = 2.61$, $SE = 0.31$) or offsets ($M = 2.50$, $SE = 0.21$) as well as the coefficient for offsetting the full supply chain ($M = 2.73$, $SE = 0.30$) were similar in strength, but influenced choice behavior less than the coefficients for Fairtrade and full reductions. The organic label ($M = 2.20$, $SE = 0.17$) had the relatively speaking smallest significant effect on the choices made. In line with expectations, the price coefficient was negative ($M = -1.33$, $SE = 0.09$), indicating that participants preferred smaller prices (all other things being equal) to higher prices. On average, the no buy option was also negative in sign ($M = -3.08$, $SE = 0.46$), indicating participants preferred choosing a coffee to not choosing a coffee across all sets. The definition of the priors (cf. chapter 3.3.4) has resulted in a noticeable impact on model precision, as the standard errors for parameters with predefined priors were comparatively smaller.

Preference heterogeneity was present for all labels. Their estimates had significant standard deviations ranging from 1.84 to 2.14 SD, except for Fairtrade with even larger deviation of $SD = 2.75$, $SE = 0.18$. The labels' influences were heterogenous across individuals, emphasizing that point-based estimates (e.g., using a multinomial model) would be insufficient to model these diverse preferences.

The second part of the table shows the added interaction terms of group membership with all eight parameter estimates to reveal possible effects correlated with the consumption of whole bean in comparison to ground coffee. None of the label estimates became significant, but both the price ($M = 0.62$, $SE = 0.09$) and no buy ($M = 2.00$ $SE = 0.49$) effects were significant and positive. The non-significant label interaction effects show that whole bean consumers did not differ significantly from ground bean consumers in their preferences for the sustainability labels. The positive price interaction term results in an overall price coefficient closer to zero for the segment, indicating that whole bean consumers were less reluctant to choose higher-priced coffee in the experiment. The positive, significant no buy estimate shows that the segment was on average more likely to choose the no buy alternative across all choice sets. This result signals that group 2 was less likely to compromise on their coffee choices if no coffee suited their preferences in a choice set.

The chosen random parameter logit model with correlated parameters also models the correlations between attribute levels. The focus of this research was not concerned with these correlations, they thus not discussed in detail in

this work. As a short summary of the findings, sustainability and decarbonization labels were significantly correlated with each other, indicating that participants whose choices were influenced by one label were also influenced by other labels. The price coefficient was not significantly correlated to any of the label coefficients.

4.5.1. Willingness to Pay Resulting from the Random Parameter Logit Model

The resulting WTP premiums based on these sustainability and decarbonization coefficients and the price coefficients are shown in table 7.

These WTP estimates show the distribution of the marginal WTP for coffee with the given label compared to filter coffee without the given label all other things equal. The interpretation of these numbers is that, on average, buyers would be equally divided between 500g (ground or whole bean) filter coffee without the attribute and an identical package with the additional attribute that costs original price plus the given amount (in €). The WTP premium of group 1 differs from group 2, as the price coefficient of group 2 was adapted based on the interaction term.

A violin plot of the WTP distributions across the full sample including a simplified boxplot is provided in figure 2. The diameter of the violin plots on the y-axis correspond to the relative frequency of individuals in the sample with the given marginal WTP shown on the x-axis scale i.e., the width of the grey area corresponds to the proportion of participants with the given WTP premium. The simplified boxplots provide a visualization of the overall median and quartiles shown in Table 7.

The plot shows that while preferences vary widely, the distributions do not indicate unexpected distribution patterns, such as distinct clusters with opposing preferences. Preference differences like these can be missed if only coefficient estimates and boxplots and not the individual parameter estimates are considered.

4.5.2. Results in Relation to the Hypotheses

In the following paragraphs, the presented results are evaluated in relation to the initial hypotheses one to four introduced in chapter 2.3.

> **H1:** On average, participants are willing to pay a premium for filter coffee that was shipped without causing net emissions.

The result of the random parameter logit shows significant estimates for emission reductions in maritime shipping ($M = 2.61$, $SE = .31$, $p<.001$), emission reductions for the full supply chain ($M = 3,67$, $SE = .35$, $p<.001$), offsets of the shipping emissions ($M = 2.50$, $SE = .21$, $p<.001$), and offsets for emissions of the full supply chain ($M = 2.73$, $SE = .30$, $p<.001$), supporting H1.

The resulting WTP derived from the individuals' level estimates and price coefficients are shown in Table 7 and Figure 2. Participants are indeed willing to pay a premium for filter coffee that was shipped without causing net emissions.

> **H2:** On average, participants value decarbonization of the full supply chain more than decarbonization of parts of the supply chain (maritime shipping only).

A two-tailed z-test showed a significant difference between the coefficients for full and partial direct reductions of emissions ($z = 2.26$, $p = .02$), but no significant difference between the two coefficients for offsets ($z = .57$, $p = .57$).

This implies that participants were valuing decarbonization of the full supply chain more than a decarbonization of the maritime shipping in the context of direct reductions but not in the context of carbon offsets. Thus, the data does not fully support H2. Counter to expectations, participants only valued decarbonization of the supply chain significantly more for direct reductions.

> **H3:** On average, participants attribute more value to direct emissions reductions compared indirect offsets.

The model coefficient for maritime shipping with direct reductions was not significantly higher than the corresponding coefficient for offsets in a one-tailed z-test, $z = .27$, $p = .78$. The coefficient of the supply chain was however significantly higher for full decarbonization compared to offsets in a second one-tailed z-test, $z = 2.05$, $p = .02$. H3 can thus not be fully supported. While participants attributed significantly less value to offsets compared to direct emission reductions for the full supply chain, they did not do so for maritime shipping.

> **H4:** On average, participants that prefer ground or whole bean filter coffee have identical preferences for decarbonization, sustainability labels and price.

The hypothesis could not be fully supported. While participants from the groups did indeed not differ significantly in their underlying preferences for decarbonization and sustainability labels, the group whole bean coffee differed significantly from group ground bean in terms of their price coefficient as shown in the significant interaction term between group and price coefficient ($M = .62$ $SE = .09$, $p<.001$). Group 1 was thus willing to pay more for coffee in general which effectively resulted in in higher WTP for sustainability labels on the package.

5. Discussion

The presented data results in a coherent description of participant behavior and stated preferences. Stated importance ratings were aligned with the behavioral measurement during the DCE. Fairtrade was both stated as one of the most important criteria when buying coffee and resulted in the biggest WTP increase. Full reductions of emissions across the supply chain were valued at similar importance as Fairtrade

Table 7: Marginal WTP for Sustainability and Decarbonization ($N = 299$)

	Full Sample			Group 1: Ground bean			Group 2: Whole bean		
Parameter	**Q1**	***Mdn***	**Q3**	**Q1**	***Mdn***	**Q3**	**Q1**	***Mdn***	**Q3**
Fairtrade label	0.94	2.77	5.62	1.04	2.27	4.29	0.81	3.32	5.94
Organic label	0.58	1.61	3.30	0.50	1.48	2.99	0.65	1.81	3.86
Offset of maritime emissions	0.46	1.79	3.88	0.65	1.78	3.19	0.43	1.92	4.29
Offset of full supply chain	0.47	1.89	4.25	0.47	1.80	3.69	0.46	2.05	4.64
Reduction of maritime emissions	0.63	1.95	3.95	0.55	1.90	3.38	0.64	2.11	4.90
Reduction of full supply chain	1.37	2.82	5.22	1.21	2.48	4.62	1.46	3.08	6.02

Note: estimates in € ; Mdn = Median, Q1 and Q3 = 1st and 3rd distribution quartiles.

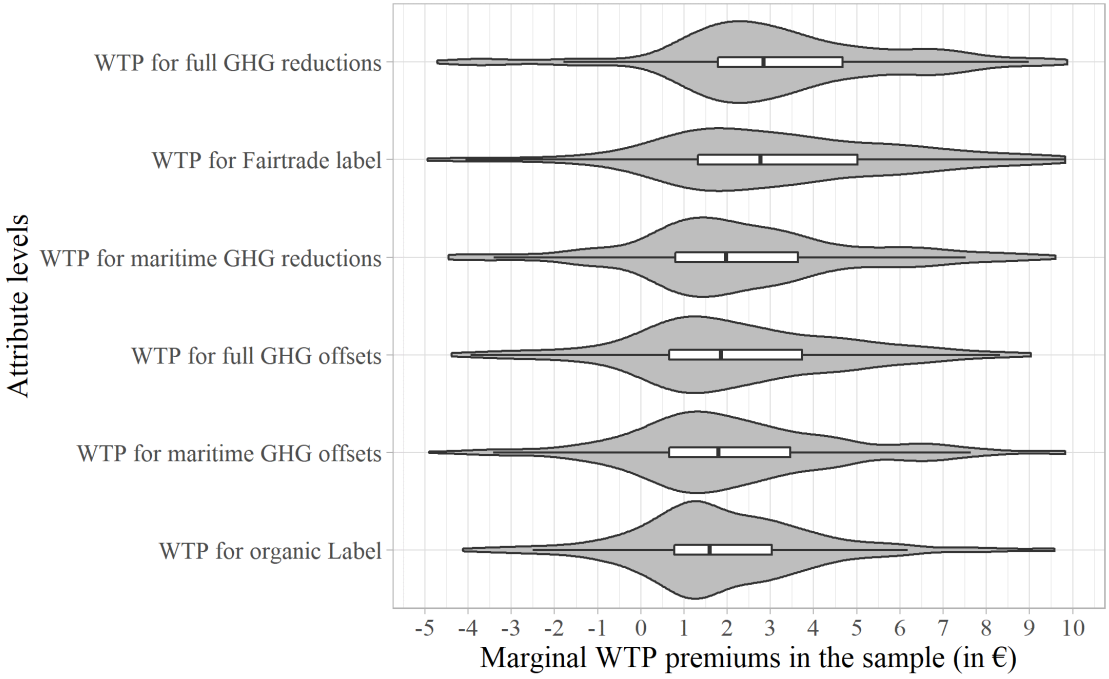


Figure 2: Distribution of Marginal WTP Premium in the Sample ($N = 299$)

certification in the DCE. The three other decarbonization labels influenced decisions less than the Fairtrade certification and full supply chain emission reductions. Nonetheless, all three decarbonization labels exerted a significant, moderate influence on decisions made. The slightly lesser valued organic label also exerted a significant positive effect on the choice. These DCE results are aligned with the stated preference ratings indicating a similar order of priority, except for the label "full decarbonization of the supply chain" which performed even better than expected in comparison to the other labels.

Stated importance of price was ranked at medium importance with a standard deviation of approximately one rank. This heterogeneity was also present in the price coefficient and is not unexpected as the in-store price comparison revealed widely varying market prices for coffee based on brands, labels, and quality. The small demographic and stated preference differences between group 1 and group 2 did not affect their modelled underlying preference for sustainability labels. However, group 2 did state a significantly lower importance of price (weak effect), an effect that was also measured and modelled in the DCE. This results in a higher WTP for coffee in general and thus effectively in higher WTP for sustainability labels on the package. While group 2 was effectively indeed willing to pay more for sustainability and decarbonization labels, they were not necessarily valuing the sustainability and decarbonization labels more compared to group 1, but instead cared less about prices for coffee in general. The average underlying value derived from sustainability labels and decarbonization seems to be rather stable across these segments. Analyzing the stated frequently bought coffees of the groups resulted in slightly

conflicting findings and did not show significant differences in reported price per kg paid between groups. However, these results relied on remembered average prices of small subsamples, thus these subtle differences revealed in the DCE might be noticeable in this less sensitive measurement.

Group 2 was more particular about their coffee choice and tended to choose the no-choice option significantly more often than group 1. Subtle group differences that might explain these results include that whole bean users were on multiple dimensions slightly more interested in product quality criteria, such as taste, smell, type of roast, and digestibility. On average, group 2 also stated to credit slightly lower importance to convenience, coherent with the fact that on average they must obviously put more effort into the preparation of their coffee. While a slightly higher age and a higher rate of students was found in group 1, there was no gender difference between groups. These results provide coherent insights into the sample structure and the stated importance ratings, but the validity of any single significant associations should not be overestimated. Most associations are highly unlikely to be random, but as the beta error was not controlled for in this context, some of the results can be random given the size of the underlying correlation matrix. No significant group differences in the importance of sustainability dimensions (e.g., organic and Fairtrade labelling) were found in the DCE.

Counter to expectations, hypotheses one and two concerned with differences in the valuation of different degrees of decarbonization could not be fully supported. The individual parameter distributions, if ordered by Median, resemble the order of valuation that was expected by the hypothesis, full decarbonization performed better than partial decarbonization and reductions performed better than offsets. These differences were however not big enough to lead to a statistically significant difference between these coefficients for full and partial offsets and for maritime shipping reductions and offsets. This could be based either on these differences being too small for the sample and design to detect or might be the result of actual indecisiveness of the consumers on these topics. Future research including stated preference data and a stronger focus on comprehension of the various forms of offsets and decarbonization could reveal possible reasons for indecisiveness. The comprehension measurements indicate that participants were worse at assigning the situational descriptions to the correct terms for situations describing offsets compared to situations describing reductions. The low overall score average is an indication for a lack of comprehension of terminology and contents and might be one explanation for the similar valuation of full and partial offsets. There were no comprehension performance differences between the two groups.

Price coefficient and coefficients for the labels resulted in plausible WTP ranges when comparing the empirical findings of the choice experiment to the prices for certified coffees as reported in the preliminary in-store price comparison. The results indicate a different priority order of WTP when comparing the results to the work of van Loo et al. (2015). While the organic label WTP premium is estimated very similarly, Fairtrade was less important to the American sample and the carbon footprint reduction pledge presented did not influence decision making significantly. Given the different content and context of the carbon label used by van Loo et al. (2015), the different and higher valuation of the carbon offset and reduction labels presented in this study seems viable. The authors did not provide a description of the label's contents beyond the vague pledge "reducing our carbon emissions", while this study's labels indicated more concrete and tangible emission reductions. Furthermore, it is not unlikely that a 2015 American sample will behave differently than a 2021 German sample in similar context. The valuation of various sustainability aspects is expected to differ across time, culture, and demographics and might have increased from 2015 to 2021 given the rise of sustainability trends.

Comparing the results to other relevant studies, at first glance the WTP premium of the whole bean segment for offsetting maritime shipping ($M = 1.92$€ per 500g) and the full supply chain ($M = 2.05$€ per 500g) look very similar to the WTP for a "carbon neutral label" reported by Birkenberg et al. (2021) if the packaging size is not considered ($M = 1.77$€ per 250g of Arabica whole beans). Fairtrade was valued less in the presented sample ($M = 3.32$€ per 500g) compared to the estimates of Birkenberg et al. ($M = 4.30$€ per 250g). These deviations might be the result of different design and sampling methods. Amongst other differences, their results rely on a sample recruited personally in German coffee shops that was informed about GHG emissions of the coffee life cycle before the experiment, possibly influencing WTP for carbon net neutrality. Other differences include the survey in a shop setting possibly priming for more expensive premium or specialty coffee of a specific brand. The public setting might also be more prone to subtle influences of social desirability compared to the online setting of this study. Furthermore, their overall sample was smaller, and the resulting model less precise in terms of standard errors compared to the model introduced. Future DCEs might consider adding varying packaging sizes as an attribute. Interestingly although possibly randomly, if the packaging size is not considered WTP for the sustainable labels seems aligned in both models. If differences in design and setting do not explain the deviations, WTP for coffee might in customer perception be less of a relative value dependent on packaging size and more of an absolute value per package of coffee bought. This could make premium segments that rely on smaller bags of coffee an attractive initial niche market for carbon reduction. Rotaris and Danielis (2011) reported a high premium for Fairtrade coffee ($M = 2.20$€ per 250g package of beans or ground coffee) in an Italian sample ($N = 135$), while Lappeman et al. (2019) reported a lower value ($M = \$1.22$ per 250g) in a South African sample ($N = 300$). Considering these international deviations, the Fairtrade WTP modelled in this study (which can be an indicator of reliability of the other labels) appears to be in a plausible range. The results of Schniederjans and Starkey (2014) and Polinori et al. (2018)

are hard to relate to this study given their different use case, design, and label descriptions leaving room for subjective interpretation of "green transportation".

Heterogeneity of WTP for the labels is evident in the significant *SD* estimates of the random parameters and visualized in Figure 2. Plotting the individual WTP premiums on a violin graph illustrated that the marginal WTP premium of most individuals was (often strongly) increased by the labels, but some individuals were also discouraged by them. This result was expected based on the findings of previous authors (Birkenberg et al., 2021; Lappeman et al., 2019; Rotaris & Danielis, 2011; van Loo et al., 2015) and is coherent with the heterogenous importance ratings and reported prices of favorite coffee. A business implication of these individual differences is possible potential for market skimming when first introducing GHG reductions (Simon & Fassnacht, 2016). Revenues generated by initially realizing higher prices with labelled coffee, specifically targeting niche segments with higher WTP, could be a vital driver to fund the necessary growth to eventually compete with container shipping at scale. Identifying and describing resulting segments of interest lies beyond the scope of this study, nonetheless whole bean coffee consumers might be a first segment to consider because of their increased overall WTP. Furthermore, the correlations between the decarbonization and sustainability labels shows interest in existing sustainability labels is related to interest in decarbonization. Whole bean customers preferring organic and Fairtrade coffee could thus be an initial audience to target.

6. Limitations

The presented study provides an applied example for pricing sustainable shipping on the use case of coffee. Multiple precautions were taken to ensure internal validity of the results, including pre-tests of the final survey, prior research to design the experiment in a suitable manner (e.g., defining meaningful price levels based on store data), consistent randomization to avoid maturation and testing effects, and limited temporary survey availability to avoid confounding effects through external events. Information exchange between participants can be assumed as minimal in the chosen online setting. The anonymous online setting minimizes investigator influence and all wordings were carefully chosen to avoid priming and unintended bias. The questionnaire is provided in Appendix A for reference and transparency. The proposed methodology closely mirrors common choice behavior in purchasing settings, however, proper framing of the proposed emission labels and a real buying context (including brands and more product variety) might moderate the observed effects and strengthen or weaken individuals' WTP.

Possible selection biases cannot be fully mitigated as a convenience sample was used. While the sample is transparently defined and rather balanced it shows tendencies towards higher education, females and individuals living in southern Germany. Future research based on representative samples is needed to establish findings that are fully representative of the German market. Measuring the influence of income and/or wealth on the price coefficient can be an interesting addition for future inquiry. While incentives could have been an additional measure to reach a bigger sample size, the statistical efficiency of the proposed modelling process led to a coherent collection of significant findings. The dropout rate of 31% might seem high for a study of the given length of around 10 minutes. A high dropout rate was however expected, as no incentives for completion were provided and the 14 consecutive choice tasks in the middle of the survey were demanding on participants. Most of the lost participants dropped out during these tasks. When estimating a viable sample in reach at not more than ∼300 participants, the choice was made to rather include more choice sets across fewer participants than to introduce fewer choice sets with possibly higher completion rates. Eventually more participants were reached, but a significant dropout and possible resulting selection effects were accepted.

The external validity of the results, i.e., whether they can be generalized, is another quality criterion to consider. Of course, the example use case of coffee cannot provide a context-independent valid estimate of WTP for sustainable (maritime) shipping, but it does provide an exemplary estimate in a rigorously controlled setting of interest. The research was not based on a representative sample of the German population, it does however provide an estimate of the size of the effects of inquiry, that needs to be measured, validated, and adapted in other contexts.

Concerns that could confound the results were monitored and thoroughly analyzed for this study. As a conclusion from the qualitative analysis of these concerns, while some concerns did arise, none threaten the overall validity of the results. Two possible factors of influence were revealed in the auxiliary parts of the survey: lacking knowledge about or comprehension of the concepts offsetting and emission reductions as well as credibility concerns regarding labels. In retrospect, an additional measure would have been useful to assess the concurrent validity of the results of the comprehension scale indicating a moderate to low average knowledge of the detailed meaning of the concepts. The categories revealed in the qualitative analysis of concerns should be considered in further research on similar labels. Future labels need to be optimized for effectiveness and easy comprehension, transparency and tangible clarity of certification. Distrust in labels, greenwashing concerns and possible doubts on the technical feasibility of emission reduction measures need to be considered and handled in this context.

Low knowledge levels of offsets and reductions were measured but could be a result of the scale used. While great care was taken in defining the scale and examples used in line with the definitions of the terms, the scale has not been validated and tested for reliability. Anyhow, low comprehension levels are not unlikely given the evidence from previous international research (Polonsky et al., 2011, 2015; Tao et al., 2021). No notable comprehension differences between the two groups were found making it unlikely that these

differences could confound results.

From a modelling perspective, outcomes of advanced statistical models can differ depending on the parameters chosen and the variables considered in the modelling process, however all criteria used for fitting and choosing the models were both transparently provided and discussed in the data analysis section (chapter 3.5) in detail to warrant objectivity and reliability of the research.

7. Conclusion

The goal of this study was to price sustainable i.e., low GHG emitting maritime shipping. Employing coffee as an example, significant willingness to pay premiums for sustainable shipping were found in the sample for fully decarbonized sustainable supply chains, but also for partial and full carbon offsets. Participants were willing to pay significantly more for full technological reductions of the supply chain emissions than for reductions of shipping emissions only and partial and full emission offsets, the latter three resulting in similar WTP. Full decarbonization of the supply chain and Fairtrade certification were associated with the highest willingness to pay premiums in the experiment. Organic certification was also perceived as a benefit and resulted in a significant, albeit smaller WTP premium. An in-store price comparison, as well as self-reported importance ratings and auxiliary survey data supported and complemented these findings.

Participants that prefer whole bean coffee to ground beans did not differ in their underlying preferences for the labels. They did however differ significantly on their price preferences and effectively *displayed a higher overall WTP for all attributes including the sustainability labels*. These differences were coherent with the group's self-reported importance ratings. Whole bean customers claimed to place more importance on quality criteria (e.g., smell and taste) and less importance on price.

Significant heterogeneity was present across all labels and groups, emphasizing the importance of segmentation and targeting for industry practitioners, as well as the importance of further inquiry into the antecedents of this heterogeneity for researchers. The benefits of employing random parameter logit models that can cope with the heterogeneity in this domain of research became evident in the modelling process. The antecedents of this heterogeneity were beyond the scope of this work, however based on existing literature known contextual and personal factors influencing decisions and attitudes in the context were provided.

The presented results extend the domain of research by providing a rigorous measurement of customer preferences and WTP for sustainable shipping and emission reductions across the supply chain. To the best of the author's knowledge, the presented discrete choice experiment is the first to assess the perceived customer value of offsets compared to direct reductions in a controlled setting for a common product use case. The study further provides an applied example for designing and performing an advanced discrete choice experiment, solely relying on openly available open-source software. The practical value of this work is to provide CargoKite and other companies and initiatives with a focus on decarbonization with reliable baseline data that shows a customer demand for carbon neutral (maritime) cargo transport.

Future research is needed to understand the underlying antecedents of the WTP heterogeneity revealed in the sample and the existing literature and validate the robustness of the results in other contexts. As significant differences across international samples were identified in literature, another interesting avenue of research could be to relate the topic to cultural dimensions and societal trends. While exceeding the scope of this study, employing segmentation approaches (e.g., cluster analysis) can be useful to better understand the individual differences that were modelled in the random parameter approach and yield useful results for stakeholders in the field of sustainable maritime shipping, e.g., by identifying and describing segments of interest and increased WTP. Furthermore, while existing research provides some interesting and actionable insights into the effective design of offset programs, labelling and related issues, industry practitioners could benefit from further research in this area.

References

Akaichi, F., de Grauw, S., Darmon, P., & Revoredo-Giha, C. (2016). Does Fair Trade Compete with Carbon Footprint and Organic Attributes in the Eyes of Consumers? Results from a Pilot Study in Scotland, The Netherlands and France. *Journal of Agricultural and Environmental Ethics*, *29*(6), 969–984. doi: 10.1007/s10806-016-9642-7

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., … Iannone, R. (2021). *rmarkdown: Dynamic Documents for R.* Retrieved from https://bookdown.org/yihui/rmarkdown/

Baranzini, A., Borzykowski, N., & Carattini, S. (2018). Carbon offsets out of the woods? Acceptability of domestic vs. international reforestation programmes in the lab. *Journal of Forest Economics*, *32*, 1–12. doi: 10.1016/j.jfe.2018.02.004

Berger, J. (2019). Signaling can increase consumers' willingness to pay for green products. Theoretical model and experimental evidence. *Journal of Consumer Behaviour*, *18*(3), 233–246. doi: 10.1002/cb.1760

Birkenberg, A., Narjes, M. E., Weinmann, B., & Birner, R. (2021). The potential of carbon neutral labeling to engage coffee consumers in climate change mitigation. *Journal of Cleaner Production*, *278*, 123621. doi: 10.1016/j.jclepro.2020.123621

Blasch, J., & Ohndorf, M. (2015). Altruism, moral norms and social approval: Joint determinants of individual offset behavior. *Ecological Economics*, *116*, 251–260. doi: 10.1016/j.ecolecon.2015.04.024

Bliemer, M. C., & Rose, J. M. (2010). Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B: Methodological*, *44*(6), 720–734. doi: 10.1016/j.trb.2009.12.004

Brazell, J. D., Diener, C. G., Karniouchina, E., Moore, W. L., Séverin, V., & Uldry, P.-F. (2006). The No-Choice Option and Dual Response Choice Designs. *Marketing Letters*, *17*(4), 255–268. Retrieved from http://www.jstor.org/stable/40216693

Bridges, J. F. P., Hauber, A. B., Marshall, D., Lloyd, A., Prosser, L. A., Regier, D. A., … Mauskopf, J. (2011). Conjoint analysis applications in health–a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, *14*(4), 403–413. Retrieved 15/07/2021, from https://www.valueinhealthjournal.com/article/S1098-3015(10)00083-5/pdf doi: 10.1016/j.jval.2010.11.013

Canzi, G., Clough, K., & Kronick, C. (2006). Joint statement on offsetting carbon emissions. *Friends of the Earth, Greenpeace and WWF-UK*. Retrieved 07/07/2021, from http://assets.wwf.org.uk/downloads/august06.pdf

Caputo, V., Nayga, R. M., & Scarpa, R. (2013). Food miles or carbon emissions? Exploring labelling preference for food transport footprint with a stated choice study. *Australian Journal of Agricultural and Resource Economics*, *57*(4), 465–482. doi: 10.1111/1467-8489.12014

Carson, R. T., Flores, N. E., Martin, K. M., & Wright, J. L. (1996). Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods. *Land Economics*, *72*(1), 80. doi: 10.2307/3147159

Carson, R. T., & Louviere, J. J. (2011). A Common Nomenclature for Stated Preference Elicitation Approaches. *Environmental and Resource Economics*, *49*(4), 539–559. doi: 10.1007/s10640-010-9450-x

Chapman, C., & Feit, E. M. (2019). *R For Marketing Research and Analytics*. Cham: Springer International Publishing. doi: 10.1007/978-3-030-14316-9

Chi, O. H., Denton, G., & Gursoy, D. (2021). Interactive effects of message framing and information content on carbon offsetting behaviors. *Tourism Management*, *83*, 104244. doi: 10.1016/j.tourman.2020.104244

Chongsuvivatwong, V. (2018). *epiDisplay: Epidemiological Data Display Package.* Retrieved from https://CRAN.R-project.org/package=epiDisplay

Corbin, J. M., & Strauss, A. L. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory / Juliet Corbin, International Institute for Qualitative Methodology, Anselm Strauss* (Fourth edition ed.). Los Angeles: SAGE.

Creutzig, F., et al. (Eds.). (2014). *Transport*. Cambridge: Cambridge University Press. Retrieved 05.07.2021, from https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_chapter8.pdf doi: 10.1017/CBO9781107415416

Croissant, Y. (2020). Estimation of Random Utility Models in R : The mlogit Package. *Journal of Statistical Software*, *95*(11). doi: 10.18637/jss.v095.i11

Croissant, Y. (2021). *dfidx: Indexed Data Frames.* Retrieved from https://CRAN.R-project.org/package=dfidx

Day, R. (2021). Carbon Offsets Are About To Become A Huge Market. *Forbes*. Retrieved 06/07/2021, from https://www.forbes.com/sites/robday/2021/04/29/carbon-offsets-are-about-to-become-a-huge-market/?sh=6e11971461ec

de Jong, M. D. T., Harkink, K. M., & Barth, S. (2018). Making Green Stuff? Effects of Corporate Greenwashing on Consumers. *Journal of business and technical communication*, *32*(1), 77–112. doi: 10.1177/1050651917729863

Deutscher Kaffeemarkt e.V. (2021). *Deutscher Kaffeemarkt 2020.* Hamburg. Retrieved 21/07/2021, from https://www.kaffeeverband.de/de/presse/deutscher-kaffeemarkt-2020-20-tassen-pro-kopf-mehr-konsumiert#

Deutscher Kaffeeverband e.V. . (2020). *Handel, Kaffeepreis und Logistik: Transport des Rohkaffees.* Hamburg. Retrieved 04.07.2021, from https://www.kaffeeverband.de/de/kaffeewissen/handel

Drichoutis, A. C., Lusk, J. L., & Pappa, V. (2016). Elicitation formats and the WTA/WTP gap: A study of climate neutral foods. *Food Policy*, *61*, 141–155. doi: 10.1016/j.foodpol.2016.03.001

EcoClipper. (2021). *EcoClipper | Emission Free Shipping and Travel.* Alkmaar, Netherlands. Retrieved 15/07/2021, from https://ecoclipper.org/

Egloff, C., Escudero, R. G., Sanders, U., Webster, J., & Zampelas, A. (2019). Sharing the Costs of IMO 2020 Across the Ecosystem. *Germany - DE*. Retrieved 27/08/2021, from https://www.bcg.com/de-de/publications/2019/sharing-costs-imo-2020-across-ecosystem

European Commission. (2021). *The organic logo: Aims of the organic logo.* Retrieved 06/07/2021, from https://ec.europa.eu/info/food-farming-fisheries/farming/organic-farming/organic-logo_de

Faber, J., & Kleijn, A. (2020). Fourth IMO Greenhouse Gas Study: Final Report. *CE Delft*. Retrieved 11.01.2020, from https://www.cedelft.eu/en/publications/2488/fourth-imo-greenhouse-gas-study

Fairtrade International. (2021a). *Absatzentwicklung und Details über Fairtrade-Kaffee in Deutschland.* Bonn, Germany. Retrieved 15/09/2021, from https://www.fairtrade-deutschland.de/produkte/kaffee/hintergrund-fairtrade-kaffee.html

Fairtrade International. (2021b). *Aims of the Fairtrade Standards.* Bonn, Germany. Retrieved 06/07/2021, from https://www.fairtrade.net/standard/aims

Felser, G. (2015). *Werbe- und Konsumentenpsychologie*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-37645-0

Fiedler, M., Frank, K., & Volland, M. (2020). *Aktuelle Entwicklungen im Fairen Handel: Umsatzentwicklungen im Geschäftsjahr 2019.* Berlin. Retrieved 04.07.2021, from https://www.forum-fairer-handel.de/fileadmin/user_upload/dateien/publikationen/materialien_des_ffh/2020_Aktuelle-Entwicklungen_im_FH_web.pdf

Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock 'n' roll / Andy Field* (4th ed. ed.). London: SAGE.

Förster, H. (2020). *Verbrauchs- und Medienanalyse - VuMA 2021.* Retrieved 04.07.2021, from https://www.vuma.de/fileadmin/user_upload/PDF/berichtsbaende/VuMA_Berichtsband_2021.pdf

Galik, C. S., & Jackson, R. B. (2009). Risks to forest carbon offset projects in a changing climate. *Forest Ecology and Management*, *257*(11), 2209–2216. doi: 10.1016/j.foreco.2009.03.017

Gustafsson, A., Herrmann, A., & Huber, F. (2007). *Conjoint Measurement: Methods and Applications* (4th ed. ed.). Berlin: Springer.

Harrell, F. E., Jr. (2021). *Hmisc: Harrell Miscellaneous.* Retrieved from https://CRAN.R-project.org/package=Hmisc

Hasselbach, J. L., & Roosen, J. (2015). Motivations behind Preferences for Local or Organic Food. *Journal of International Consumer Marketing*, *27*(4), 295–306. doi: 10.1080/08961530.2015.1022921

Hauber, A. B., González, J. M., Groothuis-Oudshoorn, C. G. M., Prior, T., Marshall, D. A., Cunningham, C., … Bridges, J. F. P. (2016). Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, *19*(4), 300–315. doi: 10.1016/j.jval.2016.04.004

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). *Applied choice analysis* (2nd edition ed.). Cambridge: Cambridge University Press. Retrieved from http://www.loc.gov/catdir/enhancements/fy1503/2014043411-b.html

Herrmann, S. (2015). *Bewertung der Lebensmittel-Labels 2015: Hintergrundbericht.* Zürich. Retrieved 15/09/2021, from https://www.wwf.ch/sites/default/files/doc-2017-10/2015-11-Hintergrundbericht-Lebensmittellabel-de.pdf

Hess, S., & Daly, A. J. (2014). *Handbook of choice modelling*. Cheltenham: Edward Elgar.

Hole, A. R. (2007). A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health economics*, *16*(8), 827–840. doi: 10.1002/hec.1197

Huber, R. A., Anderson, B., & Bernauer, T. (2018). Can social norm interventions promote voluntary pro environmental action? *Environmental Science & Policy*, *89*, 231–246. doi: 10.1016/j.envsci.2018.07.016

International Coffee Organization. (2021). *World coffee consumption.* Retrieved 07/06/2021, from http://www.ico.org/prices/new-consumption-table.pdf

International Maritime Organization. (2020a). *IMO Environment Committee approves amendments to cut ship emissions.* London. Retrieved 06.07.2021, from https://www.imo.org/en/MediaCentre/PressBriefings/pages/42-MEPC-short-term-measure.aspx

International Maritime Organization. (2020b). *IMO working group agrees further measures to cut ship emissions.* London. Retrieved 06.07.2021, from https://www.imo.org/en/MediaCentre/PressBriefings/pages/36-ISWG-GHG-7.aspx

International Maritime Organization. (2020c). *Reducing greenhouse gas emissions from ships.* London. Retrieved 19.01.2021, from https://www.imo.org/en/MediaCentre/HotTopics/Pages/Reducing-greenhouse-gas-emissions-from-ships.aspx

Johnson, R., & Orme, B. (2003). Getting the most from CBC. *Sequim: Sawtooth Software Research Paper Series, Sawtooth Software*.

Kamakura, W., Haaijer, R., & Wedel, M. (2001). The No—Choice Alternative in Conjoint Choice Experiments. *International Journal of Market Research*, *43*(1), 1–12. doi: 10.1177/147078530104300105

Klein, J. (2021). *Unsere Klimaziele: Fairer Handel mit Kaffeebohnen, fair, bio und klimaneutral: Transport per Frachtsegler.* Leipzig. Retrieved 15/07/2021, from https://cafe-chavalo.de/content/17-unsere-klimaziele

Kollmuss, A. (2010). *Handbook of carbon offset programs: Trading systems, funds, protocols and standards / Anja Kollmuss ... [et al.].* London: Earthscan.

Lappeman, J., Orpwood, T., Russell, M., Zeller, T., & Jansson, J. (2019). Personal values and willingness to pay for fair trade coffee in Cape Town, South Africa. *Journal of Cleaner Production*, *239*, 118012. doi: 10.1016/j.jclepro.2019.118012

Leifeld, P. (2013). texreg: Conversion of Statistical Model Output in R to LaTeX and HTML Tables. , *55*(8), 1–24. Retrieved from http://dx.doi.org/10.18637/jss.v055.i08

Lera-López, F., Faulin, J., & Sánchez, M. (2012). Determinants of the willingness-to-pay for reducing the environmental impacts of road transportation. *Transportation Research Part D: Transport and Environment*, *17*(3), 215–220. doi: 10.1016/j.trd.2011.11.002

List, J. A., & Gallet, C. A. (2001). What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? *Environmental and Resource Economics*, *20*(3), 241–254. doi: 10.1023/A:1012791822804

Liu, M., & Wronski, L. (2018). Examining Completion Rates in Web Surveys via Over 25,000 Real-World Surveys. *Social Science Computer Review*, *36*(1), 116–124. doi: 10.1177/0894439317695581

Loomis, J. B. (2014). 2013 WAEA Keynote Address: Strategies for Overcoming Hypothetical Bias in Stated Preference Surveys. *Journal of Agricultural and Resource Economics*, *39*(1), 34–46. Retrieved from http://www.jstor.org/stable/44131313

Lusk, J. L. (2003). Effects of Cheap Talk on Consumer Willingness–to–Pay for Golden Rice. *American Journal of Agricultural Economics*, *85*(4), 840–856. doi: 10.1111/1467-8276.00492

Meunier, C., Stoll, J., & Schoen, L. (2019). *Freiwillige CO2-Kompensation.* Dessau-Roßlau. Retrieved 07/07/2021, from https://www.umweltbundesamt.de/themen/freiwillige-co2-kompensation

Murphy, J. J., Allen, P. G., Stevens, T. H., & Weatherhead, D. (2005). A Meta-analysis of Hypothetical Bias in Stated Preference Valuation. *Environmental and Resource Economics*, *30*(3), 313–325. doi: 10.1007/s10640-004-3332-z

Norwood, F. B., & Lusk, J. L. (2011). Social Desirability Bias in Real, Hypothetical, and Inferred Valuation Experiments. *American Journal of Agricultural Economics*, *93*(2), 528–534. doi: 10.1093/ajae/aaq142

Oxford University Press. (2021). *Greenwashing - Definition.* Oxford. Retrieved 26/07/2021, from https://www.oxfordlearnersdictionaries.com/definition/english/greenwash

Parker, J. R., & Schrift, R. Y. (2011). Rejectable Choice Sets: How Seemingly Irrelevant No-Choice Options Affect Consumer Decision Processes. *Journal of Marketing Research*, *48*(5), 840–854. doi: 10.1509/jmkr.48.5.840

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the Correct Statistical Test for the Equality of Regression Coefficients. *Criminology*, *36*(4), 859–866. doi: 10.1111/j.1745-9125.1998.tb01268.x

Pitt, A. (2017). *IMO 2020 Regulation Could Cost Shippers Extra US$60 Billion a year.* Singapore. Retrieved 14.07.2021, from https://www.woodmac.com/press-releases/imo-2020-regulation-shippers-us$60-billionyear/

Pleeging, E., van Exel, J., Burger, M. J., & Stavropoulos, S. (2021). Hope for the future and willingness to pay for sustainable energy. *Ecological Economics*, *181*, 106900. doi: 10.1016/j.ecolecon.2020.106900

Polinori, P., Marcucci, E., Gatta, V., Bigerna, S., Bollino, C. A., & Micheli, S. (2018). Eco-Labeling And Sustainable Urban Freight Transport: How Much Are People Willing To Pay For Green Logistics ? *Rivista Internazionale di Economia dei Transporti / International Journal of Transport Economics*, *XLV*, 631–658. doi: 10.19272/201806704006

Polonsky, M. J., Garma, R., & Landreth Grau, S. (2011). Western consumers' understanding of carbon offsets and its relationship to behavior. *Asia Pacific Journal of Marketing and Logistics*, *23*(5), 583–603. doi: 10.1108/13555851111183048

Polonsky, M. J., Grau, S. L., & Garma, R. (2015). Exploring US Consumers Understanding of Carbon Offsets. In L. Robinson (Ed.), *Proceedings of the 2009 Academy of Marketing Science (AMS) Annual Conference* (pp. 14–18). Cham: Springer International Publishing. doi: 10.1007/978-3-319-10864-3{\textunderscore}10

Psaraftis, H. N. (2019). *Sustainable Shipping*. Cham: Springer International Publishing. doi: 10.1007/978-3-030-04330-8

Qualtrics. (2021). *Qualtrics XM.* Provo, Utah, USA: Author. Retrieved 17.02.2021, from https://www.qualtrics.com

R Core Team. (2021). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rao, V. R. (2014). *Applied Conjoint Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-87753-0

Raza, Z. (2020). Effects of regulation-driven green innovations on short sea shipping's environmental and economic performance. *Transportation Research Part D: Transport and Environment*, *84*, 102340. doi: 10.1016/j.trd.2020.102340

Ritchie, B. W., Kemperman, A., & Dolnicar, S. (2021). Which types of product attributes lead to aviation voluntary carbon offsetting among air passengers? *Tourism Management*, *85*, 104276. doi: 10.1016/j.tourman.2020.104276

Rose, J. M., & Bliemer, M. C. J. (2013). Sample size requirements for stated choice experiments. *Transportation*, *40*(5), 1021–1041. doi: 10.1007/s11116-013-9451-z

Rotaris, L., & Danielis, R. (2011). Willingness to Pay for Fair Trade Coffee: A Conjoint Analysis Experiment with Italian Consumers. *Journal of Agricultural & Food Industrial Organization*, *9*(1). doi: 10.2202/1542-0485.1305

RStudio Team. (2021). *RStudio: Integrated Development Environment for R.* Boston, MA: RStudio, PBC. Retrieved from http://www.rstudio.com/

Schmidt, J., & Bijmolt, T. H. A. (2020). Accurately measuring willingness to pay for consumer goods: a meta-analysis of the hypothetical bias. *Journal of the Academy of Marketing Science*, *48*(3), 499–518. doi: 10.1007/s11747-019-00666-6

Schniederjans, D. G., & Starkey, C. M. (2014). Intention and willingness to pay for green freight transportation: An empirical examination. *Transportation Research Part D: Transport and Environment*, *31*, 116–125. doi: 10.1016/j.trd.2014.05.024

Schwirplies, C., Dütschke, E., Schleich, J., & Ziegler, A. (2017). *Consumers' willingness to offset their CO2 emissions from traveling: A discrete choice analysis of framing and provider contributions: Working Papers "Sustainability and Innovation* (No. S05/2017). Retrieved from https://ideas.repec.org/p/zbw/fisisi/s052017.html

Schwirplies, C., Dütschke, E., Schleich, J., & Ziegler, A. (2019). The willingness to offset CO2 emissions from traveling: Findings from discrete choice experiments with different framings. *Ecological Economics*, *165*, 106384. doi: 10.1016/j.ecolecon.2019.106384

Serra, P., & Fancello, G. (2020). Towards the IMO's GHG Goals: A Critical Overview of the Perspectives and Challenges of the Main Options for Decarbonizing International Shipping. *Sustainability*, *12*(8), 3220. doi: 10.3390/su12083220

Simon, H., & Fassnacht, M. (2016). *Preismanagement*. Wiesbaden: Springer Fachmedien Wiesbaden. doi: 10.1007/978-3-658-11871-6

Singer, E., & Ye, C. (2013). The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 112–141. doi: 10.1177/0002716212458082

Sonnenschein, J., & Smedby, N. (2019). Designing air ticket taxes for climate change mitigation: insights from a Swedish valuation study. *Climate Policy*, *19*(5), 651–663. doi: 10.1080/14693062.2018.1547678

Steiner, B. E., Peschel, A. O., & Grebitus, C. (2017). Multi-Product Category Choices Labeled for Ecological Footprints: Exploring Psychographics and Evolved Psychological Biases for Characterizing Latent Consumer Classes. *Ecological Economics*, *140*, 251–264. doi: 10.1016/j.ecolecon.2017.05.009

Streimikiene, Balezentis, Alisauskaite-Seskiene, Stankuniene, & Simanaviciene. (2019). A Review of Willingness to Pay Studies for Climate Change Mitigation in the Energy Sector. *Energies*, *12*(8), 1481. doi: 10.3390/en12081481

Tao, Y., Duan, M., & Deng, Z. (2021). Using an extended theory of planned behaviour to explain willingness towards voluntary carbon offsetting among Chinese consumers. *Ecological Economics*, *185*, 107068. doi: 10.1016/j.ecolecon.2021.107068

Traets, F., Sanchez, D. G., & Vandebroek, M. (2020). Generating Optimal Designs for Discrete Choice Experiments in R : The idefix Package. *Journal of Statistical Software*, *96*(3). doi: 10.18637/jss.v096.i03

Train, K. E. (2009). *Discrete Choice Methods with Simulation* (Second Edition ed.). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511805271

United Nations. (2016). *Maritime Transport Is 'Backbone of Global Trade and the Global Economy', Says Secretary-General Ban Ki-moon in Message for International Day*. London. Retrieved 06.07.2021, from https://www.un.org/press/en/2016/sgsm18129.doc.htm

United Nations Treaty Collection. (2015). *The Paris Agreement*. United Nations Framework Convention on Climate Change. Retrieved 05.07.2021, from https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

Upham, P., Dendler, L., & Bleda, M. (2011). Carbon labelling of grocery products: public perceptions and potential emissions reductions. *Journal of Cleaner Production*, *19*(4), 348–355. doi: 10.1016/j.jclepro.2010.05.014

Usva, K., Sinkko, T., Silvenius, F., Riipi, I., & Heusala, H. (2020). Carbon and water footprint of coffee consumed in Finland—life cycle assessment. *The International Journal of Life Cycle Assessment*, *25*(10), 1976–1990. doi: 10.1007/s11367-020-01799-5

van Loo, E. J., Caputo, V., Nayga, R. M., Seo, H.-S., Zhang, B., & Verbeke, W. (2015). Sustainability labels on coffee: Consumer preferences, willingness-to-pay and visual attention to attributes. *Ecological Economics*, *118*, 215–225. doi: 10.1016/j.ecolecon.2015.07.011

Waring, E., Quinn, M., McNamara, A., Arino de la Rubia, E., Zhu, H., & Ellis, S. (2021). *skimr: Compact and Flexible Summaries of Data.* Retrieved from https://CRAN.R-project.org/package=skimr

Wickham, H. (2021). *tidyr: Tidy Messy Data.* RStudio, PBC. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Romain, F., Henry, L., & Müller, K. (2021a). *dplyr: A Grammar of Data Manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., Romain, F., Henry, L., & Müller, K. (2021b). *ggplot2: Elegant Graphics for Data Analysis.* Retrieved from https://CRAN.R-project.org/package=dplyr

Willer, H., Trávníček, J., Meier, C., & Schlatter, B. (2021). *The World of Organic Agriculture: Statistics & Emerging Trends 2021* (22nd ed.). Bonn. Retrieved 28.09.2021, from https://www.fibl.org/fileadmin/documents/shop/1150-organic-world-2021.pdf

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R Markdown: The definitive guide*. Boca Raton: Taylor & Francis CRC Press.

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R Markdown cookbook* (1st ed.). Boca Raton: Chapman & Hall/CRC.

Junior Management Science 7(3) (2022) 569-603

Junior Management Science

journal homepage: www.jums.academy

The Variety of CSR Disclosure and its Relationship with the Underlying Performance: A Textual Analysis

Niklas Winterberg

WHU – Otto Beisheim School of Management

Abstract

This study examines the relationship between corporate social responsibility (CSR) disclosure and the underlying CSR performance. The linguistic features of integrated and stand-alone sustainability reports from companies listed in the STOXX Europe 600 between 2010 and 2018 are investigated using computer-based textual analysis. The observed textual features are as follows: length, numeric content, horizon content, target orientation, readability, tone, topic-specific disclosure, and the number of topics covered. Additional variables include whether the report is following the framework of the Global Reporting Initiative (GRI), whether the CSR information is integrated into the annual report or prepared as a stand-alone report, and whether the company is defined as an early or late adopter. Concerning the relationship between the textual characteristics and CSR performance, the results support the hypotheses that length, target orientation, and the number of GRI topics covered are positively correlated with the performance. Concerning topic-specific disclosure, only the environmental and social dimensions are positively correlated with the corresponding performance. The results also reveal that companies with superior CSR performance tend to publish stand-alone reports under the GRI framework and started reporting before the announcement of the EU Directive in 2014.

Keywords: Corporate social responsibility; sustainability; sustainability reporting; textual analysis; GRI.

1. Introduction

In recent decades, companies have experienced an increasing demand from their stakeholders and shareholders to be informed about the company's impact on sustainability (Nazari, Hrazdil, & Mahmoudian, 2017). As climate change becomes an increasingly important issue, the public strives to understand how global companies are trying to solve this problem. Nonetheless, the demand is not only driven by the growing awareness of climate change. Human rights violations and other corporate scandals, such as accusations of child labor against Apple, Samsung, and other technology companies, have also led to external pressure for sustainability reporting (Wakefield, 2016). However, internal benefits equally caused that companies started reporting on their sustainability impact. These benefits are, for instance, that companies understand the underlying risks and improve efficiency of their operational processes (Global Reporting Initiative, 2016c). Consequently, a growing trend towards sustainability reporting has been observed in recent years (KPMG, 2017).

Due to this global trend, governments started to regulate sustainability reporting similarly to financial reporting. With reference to the latter, the European Union (EU) decided in 2002 that all listed companies in the EU must adopt the International Financial Reporting Standards (IFRS) (European Commission, n.da). As a sustainability counterpart to this regulation, the EU introduced the Directive 2014/95/EU. This regulation mandates that large firms have to prepare reports on their operations' environmental and social impacts (European Commission, n.db). A similar regulation has already come into force earlier in other countries, such as France, Denmark, and the United Kingdom (UK). These laws also mandate that companies disclose certain sustainability information (Fiechter, Hitz, & Lehmann, 2019; Hummel & Rötzel, 2019). In contrast to financial reporting, the EU Directive does not require companies to have their reports audited in detail or use a specific framework (European Commission, n.db).

Various frameworks and guidelines that support compa-

DOI: https://doi.org/10.5282/jums/v7i3pp569-603

nies during the preparation process of their sustainability reports emerged. These are designed to ensure that the reports' format and content are consistent and comparable among companies. According to a survey by KPMG (2017), the GRI framework is the most widely accepted framework for sustainability reporting. Nonetheless, there are also many other guidelines, such as the United Nations Global Compact (UNGC), the ISO 26000, and the OECD Guidelines for Multinational Enterprises.

The combination of a weak legal situation and a large number of guidelines allows companies to design their reports very individually. The variety of sustainability reports is further reinforced by the fact that companies are not required to disclose this information together with the annual report as long as it is "made publicly available within a reasonable period of time [...] after the balance sheet date, on the undertaking's website" (European Union, 2014, p. 5). Thus, they can independently decide whether they publish this information within their annual report or separately in a stand-alone sustainability report. As a result, this leads to a high diversity of sustainability reports.

The diversity of sustainability reports offers numerous research opportunities and is the fundamental motivation for this study. This study examines the differences between sustainability reports and how the corresponding textual characteristics are related to the underlying sustainability performance. For this purpose, a textual analysis based on statistical software is performed. This method allows to objectively and automatically retrieve textual characteristics from more than 2,500 reports. The observed textual characteristics are as follows: report length, readability, tone, horizon content, numeric content, target orientation, topic-specific CSR disclosure, and the number of GRI items covered. In addition, other essential variables include the use of the GRI framework, the report type, and whether the company is defined as an early or late adopter.

The first part deals with the differences in reports from specific groupings and with the textual development over time. Regarding the latter analysis, one could expect that the announcement of the EU Directive or its entry into force three years later might affect the reports' linguistic design. In addition to these events, there might be some other changes due to the general sustainability reporting trend. Moreover, reports from early and late adopters are compared. Similar to financial reporting, some companies have started to report on their sustainability responsibilities on their own initiative, while other companies have been moved to do so by regulatory pressure. This is the fundamental difference between early and late adopters. The study aims to examine the differences between the reports from both adopter types and to explore the underlying reasons. In addition, there are two other groups: integrated reports and stand-alone reports. The former are annual reports that include sustainability information, while the latter are separate reports that contain no further information.

The second part of the thesis deals with the relationship between sustainability disclosure and the corresponding performance. The motivation is to determine whether companies with a superior sustainability performance design their reports differently from companies with below-average performance. A regression is performed to investigate this relationship. The dependent variable is an economic, social, and governance (ESG) score, while the textual features and the other variables mentioned above are the independent variables. Moreover, additional analyses as well as different robustness tests are performed to confirm the key findings.

The remainder of this paper is structured as follows: Chapter two, *Corporate Social Responsibility Disclosure,* defines the concept of sustainability reporting and explains the EU Directive, which was announced in 2014. Afterwards, chapter three, *Frameworks & Guidelines,* introduces the different sustainability reporting frameworks. Thereby, it describes the GRI and the International Integrated Reporting Council (IIRC) as well as the corresponding frameworks. Chapter four, *Methodology,* explains how the study, in particular the textual analysis, was conducted. Afterwards, chapter five, *Literature Review & Hypothesis Development,* summarizes the current literature's findings and creates several hypotheses for the relationship between sustainability performance and the textual features. In the next step, chapter six, *Descriptive Statistics,* investigates the differences between integrated and stand-alone sustainability reports, the differences between reports from early and late adopters, and the textual development of sustainability reports over time. Furthermore, chapter seven, *Implications for CSR Performance,* contains the regression structure and tests the corresponding assumptions. Chapter eight, *Results,* includes the final results of the regression and relates to the formulated hypotheses. Moreover, it also elaborates on different robustness tests and additional analyses and includes managerial and theoretical implications. Finally, the thesis comes to a conclusion by addressing the limitations and summarizing the insights of this paper.

2. Corporate Social Responsibility Disclosure

2.1. Definition of Sustainability Reporting

The concept of sustainability is manifold and comprises various facets. In the current literature, it is often characterized as a three-pillar conception and includes the following three aspects: "economic", "social", and "environmental" (Purvis, Mao, & Robinson, 2019). In 1987, the United Nations Brundtland Commission defined sustainability as "meeting the needs of the present without comprising the ability of future generations to meet their own needs" (United Nations, 1987, p. 16). Thus, a major aspect of the concept is the long-term orientation and that economic, social, and ecological components can be beneficially combined. Transferred to a business setting, it relates to the companies' long-term survival by considering and combining the three perspectives (INTOSAI, 2013). The terms "sustainability", "corporate social responsibility", "corporate responsibility" (CR), or "economic, social, and governance" are often used

interchangeably to describe the same concept. Therefore, in this paper, they will not be further differentiated and utilized as synonyms (Ioannou & Serafeim, 2017).

Since traditional reporting and accounting is mainly based on financial disclosure, it is often criticized that this reporting type does not accurately represent the overall company due to missing information on environmental and social aspects (Aureli, Medei, Supino, & Travaglini, 2016). Internal as well as external stakeholders have pressured companies to start reporting on their sustainability responsibilities (Ballou, Heitger, Landes, & Adams, 2006). These reports are often declared as "Sustainability Report", "ESG Report", or "CSR Report". Often, non-financial information is also integrated into the annual report instead of being published in a separate report. In general, CSR or sustainability reporting can be defined as providing "nonfinancial information to key stakeholders, those people affected by a company's actions, on the company's operational, social, and environmental activities and its abilities to deal with related risks" (Ballou et al., 2006, p. 66). Thereby, it displays the linkage between the strategy of the company and its commitment to sustainable development. It also improves the companies' understanding of their sustainability performance and supports the company in measuring and communicating it (Global Reporting Initiative, n.db).

The development of sustainability reporting has started in the late 1980s when the first voluntary environmental reports have been disclosed. Large companies with highly pollutive operations were pressured by non-governmental organizations (NGOs) to report on these issues. In the following years, the reporting scope was widened, and CSR reporting was developed. Major reasons for this were concerns about human rights and labor conditions within the organizational structure and among the supply chain (INTOSAI, 2013). Another milestone of sustainability reporting was in 1998 when Elkington (1998) introduced the triple-bottom-line. This concept is based on the traditional bottom-line perspective and focuses not only on the economic value of a company but also on social and environmental values. A global trend has emerged during the recent decade with more and more companies starting to report on their sustainability activities. In 2011, around 44% of the G250 companies, which are the largest 250 corporations worldwide by revenue, disclosed CSR information within their annual reports or stand-alone reports. Six years later, the share of CSR reporting companies has increased to around 80% (KPMG, 2017).

Nowadays, financial investors are not only interested in financial performance but also in the management of sustainability issues. This trend can be seen in the global growth of sustainable investment funds or the introduction of specialized investment rating systems, such as the Dow Jones Sustainability Index (Pencle & Mălăescu, 2016). Hence, this trend underlines that social and environmental information is essential for various stakeholders' decision-making processes (Cho, Roberts, & Patten, 2010). The underlying reasons are not only moral or ethical issues but also the financial implications of sustainability issues. Enormous carbon dioxide (CO_2) emissions, for instance, might have a financial impact since the company has to buy emissions certificates or suffers a drop in sales due to a harmed company image (INTOSAI, 2013). Therefore, companies do not only have to follow the approach of maximizing shareholder value but also have to consider the interests of other key stakeholders (Ballou et al., 2006). In turn, this explains why an increasing number of companies have started reporting on their sustainability issues in recent years.

However, companies are not only extrinsically but also intrinsically motivated to pursue this reporting approach. On the one hand, for instance, sustainability reporting allows companies to better understand the underlying risks. On the other hand, employees might appreciate the efforts leading to higher job satisfaction. In turn, this might positively affect productivity and reduce employee turnover. Moreover, even if the reporting process requires time and effort, companies might benefit from it through resource and financial savings in the medium- and long-run. These, for instance, can be caused by reductions in energy or resource consumption (INTOSAI, 2013).

In contrast to financial reporting, sustainability reports do not only address shareholders but a wide range of stakeholders, for instance, local communities, environmental organizations, and the general public. These reports often cover the companies' philanthropic, environmental, operational, social, and economic objectives and how they manage the related risks (Ballou et al., 2006). However, the content is often very industry-specific. While automotive companies tend to focus on emissions and fuel consumption, textile companies tend to elaborate on topics like child labor and human rights (Liew, Adhitya, & Srinivasan, 2014).

All in all, sustainability reporting has become increasingly important in recent years and is used by companies to inform their stakeholders about their social and environmental responsibilities.

2.2. EU Directive

While financial reporting is rigorously controlled and regulated, the regulatory framework for disclosing non-financial information has been relatively weak. In 2013, the EU proposed the Directive 2014/95/EU, also called the non-financial reporting directive (NFRD). Thereby, the EU intended to improve the transparency of large European companies by mandating the disclosure of non-financial information. One year later, the directive was approved by the Council of the EU, and member states had to implement it within the following two years. The regulation has been effective since 2017. Thus, the first mandatory reports were published in 2018, covering the financial year 2017-2018 (European Commission, n.db; Global Reporting Initiative, n.df). Thereby, the directive led to a shift from voluntary to mandatory disclosure of non-financial information.

European companies of "public-interest", which employ more than 500 employees, must comply with the directive. In total, this amounts to approximately 6,000 entities within

the EU and includes banks, listed companies, insurance companies, or companies that are declared as "public-interest" entities by the authorities of the individual country (Global Reporting Initiative, n.df).

These companies must report on the following four buckets: environmental protection, anti-corruption, human rights, and social responsibility. While environmental protection comprises aspects like the usage of renewable energy and air pollution, social responsibility, for example, is related to working conditions and labor unions. Additionally, another compulsory part is the description of the business model, as well as a description of the diversity policies established for supervisory and management bodies. Companies must also explain the underlying risks and outcomes of policies that are implemented to tackle the four buckets.

However, if a company does not apply policies for one of those buckets, it is required to explain the reasons. Thereby, reporting flexibility is granted by the EU. Companies can either describe the applied policies or explain the underlying reasons for the non-existence of such policies. Furthermore, the directive neither includes an obligation to utilize a recognized reporting framework nor requires an audit company to accurately verify the information. Nevertheless, it encourages the use of frameworks, such as the UNGC or the GRI's Sustainability Reporting Guidelines (Global Reporting Initiative, n.df). With reference to assurance, it is only required that an independent audit firm checks the existence of non-financial information (European Union, 2014). Thus, the EU Directive does not impose any clear restrictions on the report design.

To sum up, the EU introduced the regulatory framework to enhance the transparency of major European enterprises. Thereby, it aims to improve the environmental and social performance of these companies. The EU also expects a positive long-term impact on economic growth and employment. Another objective is to increase the number of reporting companies and improve sustainability reports' information quality. Currently, the European Commission (EC) reviews the NFRD, which was decided at the end of 2019. In this way, the EC aims to continuously improve the directive and strengthen the sustainable development within the EU (Global Reporting Initiative, n.df). Even if the regulation was enormously criticized for its flexibility, it is considered as a milestone for sustainable development (Mittelbach-Hoermanseder, Hummel, & Rammerstorfer, 2019).

3. Frameworks & Guidelines

3.1. Overview of Frameworks & Guidelines

Referring to the regulatory framework for financial reporting in the EU, listed companies have to prepare their financial reports in accordance with the IFRS (European Commission, n.da). In contrast, there is no regulatory requirement for the use of a guiding framework referring to sustainability reporting (Nazari et al., 2017). As described above, even the implementation of the EU Directive did not lead to a mandatory reporting framework. As a result, companies can individually decide on their reporting methodology, which resulted in various reporting practices. Comparing the reports from different companies, they vary, for instance, regarding the format and the utilized performance indicators (Nazari et al., 2017). While some companies, for example, incorporate their sustainability information into their annual report, other companies publish biannually stand-alone reports (INTOSAI, 2013). In turn, according to a study by PwC (2014), the majority of financial investors worldwide are not satisfied with the heterogeneity of the current sustainability reporting practices.

Hence, numerous international and national initiatives came into existence, aiming to standardize and harmonize reporting practices. Depending on the industry, company, or strategic orientation, the different guidelines serve as a framework for a systematic reporting approach. The most commonly used framework is provided by the GRI and will be explained in the following chapter. Other frameworks include the OECD Guidelines for Multinational Enterprises or the UNGC initiated by the United Nations (UN). The latter comprises ten social and ecological standards, and around 12,000 companies from 170 countries are committed to the principles. Thereby, these firms have to publish an annual report called Communication on Progress (COP). On top of that, national guidelines like the Connected Reporting in the UK or the Sustainability Code in Germany are often applied by small enterprises with a short value chain or scarce financial resources (BMAS, n.d; INTOSAI, 2013). Since some companies integrate their sustainability information into their annual report, the IIRC provides guidelines for these report types.

3.2. Global Reporting Initiative

3.2.1. Overview & History

The Global Reporting Initiative is an independent international organization founded in 1997 in Boston and headquartered in Amsterdam (Global Reporting Initiative, n.da). It was initiated as a shared initiative of the UN Environmental Program and the Coalition for Environmentally Responsible Economies, a NGO based in the United States (US) (Clarkson, Li, Richardson, & Vasvari, 2008). Its global presence is supported through regional hubs in numerous countries, such as Colombia and India (Global Reporting Initiative, n.da). Furthermore, it is a non-profit foundation funded by various sources like partner organizations, businesses, and governments (Global Reporting Initiative, n.dc). The organization aims to improve the reporting quality of businesses worldwide. For this purpose, it supports governments and companies in understanding and communicating their sustainability impact. This attitude is reflected in their mission "to empower decisions that create social, environmental and economic benefits for everyone" (Global Reporting Initiative, n.da).

The organization's core product is the GRI sustainability reporting framework, which is called the GRI Standards. The

first version was released in 2000. It is the first framework for sustainability reporting and is also the most widely adopted one. Since the first release, the guidelines have been further developed and improved and rely on global best practices (Global Reporting Initiative, n.da). In 2017, 75% of all reports published by G250 companies were in line with the GRI framework (KPMG, 2017). Furthermore, the guidelines are universally applicable and can be applied by any organization independent of its size, location, sector, or whether it is a public or a private company. Thereby, GRI aims to ensure that stakeholders receive comparable data from different companies (Global Reporting Initiative, n.db).

The guidelines are based on a multi-stakeholder approach by bringing different stakeholders together to participate in a dialogue and jointly decide on changes and improvements. This approach should guarantee that different perspectives are considered. Moreover, during the continuous development and improvement process, a consensus-seeking approach is utilized. This procedure ensures that the interests of various stakeholders, for instance, businesses, governments, and labor unions, are taken into account (Global Reporting Initiative, n.dc).

Apart from developing its reporting framework, GRI also actively promotes the implementation at the company level. In the course of this, the organization advises market regulators, stock exchanges, and governments during the development process of sustainability policies. The collaborative approach aims to create a beneficial environment for CSR reporting (Global Reporting Initiative, n.da). As a result, GRI is referenced in the sustainability policies of 35 countries. Moreover, the organization also collaborates with other issuers of reporting guidelines like the UNGC and the International Organization for Standardization (ISO). In cooperation, GRI published guidelines on how to combine different frameworks to harmonize the reporting process (Global Reporting Initiative, n.dd). Moreover, it cooperates with these organizations to avoid content duplication (Global Reporting Initiative, n.de).

3.2.2. Structure of GRI Standards

In 2016, GRI published the latest version of the GRI Standards, which replaced the predecessor framework G4. Apart from content clarifications and the usage of a more straightforward language, the overall structure was changed. Thereby, the content has been restructured into a modular structure consisting of six interrelated sets of standards. Three of these sets are universal standards, which are applied by every reporting company. Additionally, three topic-specific buckets cover economic, environmental, and social issues. The modular structure grants higher flexibility since individual standards within a set can be modified without revising the whole set. Hence, the framework is more responsive to changes and can be regularly updated (GRI Secretariat, 2016).

In general, each standard includes the following three aspects: requirements, recommendations, and guidance. Requirements are all mandatory instructions. These have to be fulfilled by the reporting company to prepare a sustainability report that is in accordance with the GRI Standards. While requirements are labeled with the modal verb “shall”, recommendations are indicated with the modal verb “should”. The latter one represents measures which are advised but not mandatory. Lastly, the guidance part provides explanations and background information to support businesses in understanding the different guidelines. Thereby, this section also includes examples that act as an orientation for the reporting company (Global Reporting Initiative, 2016a).

The first universal standard is “GRI 101 – Foundation”. This standard acts as a starting point for the reporting process and includes the ten reporting principles. These principles cover topics like reliability, comparability, and completeness, ensuring reporting quality and sufficient reporting content. Moreover, it explains how to use the various standards and also how to reference them. In the same part, companies have to conduct the materiality assessment to identify the topics with the most significant impact on stakeholders (GRI Secretariat, 2016). Furthermore, GRI 101 also includes the requirements for publishing a report, which is in accordance with the GRI Standards, or for publishing a report, which contains only selected standards (Global Reporting Initiative, 2016a).

In the second universal standard, “GRI 102 – General Disclosures”, businesses have to report contextual information. This standard should ensure that stakeholders are able to understand the business model as well as the environment of the reporting company. Hence, this set covers topics like the organizational profile, the company’s strategy, and the reporting practice. In this standard, companies can decide on reporting in accordance with the GRI Standards core or comprehensive option. For the latter option, companies have to report on all disclosure items. In contrast, they only have to report fewer mandatory items in the core option, as it will be described later (Global Reporting Initiative, 2016b).

The third universal standard, “GRI 103 – Management Approach”, has to be utilized with the topic-specific buckets. Thereby, it should explain the reasons for the topic’s materiality and the organization’s approach to tackle this issue. If the organization has not implemented a particular management approach, it has to explain the underlying reasons. This is in line with the “report or explain” approach of GRI. Moreover, the company is required to evaluate the management approach with regard to its effectiveness (Global Reporting Initiative, 2016c).

The three topic-specific buckets are the economic, social, and environmental standards series. While the economic standards comprise topics like procurement practices (GRI-204) and anti-corruption issues (GRI-205), the environmental standards focus on topics like water usage (GRI-303), biodiversity (GRI-304), and waste management (GRI-306). With regard to the social standards, the content is, for instance, related to occupational health and safety (GRI-403), child labor (GRI-408), and customer privacy (GRI-418). In total, all three topic-specific buckets include 34 individual standards (GRI Secretariat, 2016). However, the range of

these topics is not exhaustive. If a particular material topic does not accurately match one of the standards, the company still has to report "other appropriate disclosures" (Global Reporting Initiative, 2016a, p. 19).

Moreover, there are two different approaches on how to use the GRI Standards. Companies can utilize the guidelines as a set, or they can only use selected items to report on specific topics. Regardless of the reporting method, businesses can individually decide on publishing a stand-alone sustainability report or an integrated report. Moreover, they can also reference information that is disclosed in other electronic or paper-based locations.

The first approach is the extensive version to present a complete picture of the company's sustainability responsibilities and the underlying management approach. Following this approach, companies prepare a report *in accordance with the GRI Standards*. Each reporting company, which wants to pursue this reporting approach, must publish a GRI index that contains the items on which the company has reported. Furthermore, these companies need to conduct a materiality assessment and report on every topic with a material impact on sustainability (Global Reporting Initiative, 2016a).

This approach is sub-divided into a comprehensive and a core option. The latter one is a downsized version, including the minimum amount of information to comprehend the organization and its impact on stakeholders. The comprehensive option is an extended version that contains additional information on the company's strategy, governance, and ethics. In addition, the company has to report on every topic-specific item within a material topic. Referring to the topic-specific item "GRI 303 – Water", for instance, the company would have to report on all three disclosure items (303-1; 303-2; 303-3) to fulfill the requirements of the comprehensive option. The two options refer only to the application of the GRI Standards but not to the reporting quality (Global Reporting Initiative, 2016a).

The second approach includes the utilization of only specific standards, if the company does not want to report on all of its material topics. This is called the *GRI-referenced* claim and allows businesses to report on specific information without providing a complete overview of its material topics (Global Reporting Initiative, 2016a).

3.3. Integrated Reporting

In 2002, a Danish biotechnology company called Novozymes published the first annual report that included information on environmental and social aspects. The former CEO of the company, Steen Riisgaard, stated that Novozymes decided to combine the information in one report since "various stakeholders [were] asking for a wider overview of the business" (Eccles, Krzus, & Solano, 2019, p. 2). Nowadays, these reports are called integrated reports since traditional financial reporting is combined with reporting on sustainability responsibilities. In the following years, more and more companies, such as BASF in 2007, have adopted this reporting type and issued integrated reports. However, these firms did not apply a common framework during the preparation process. Hence, the comparability of these reports was relatively low (Eccles et al., 2019).

To ensure uniform and comparable integrated reports, the Prince's Accounting for Sustainability (A4S) Project and GRI jointly launched the International Reporting Committee in 2010. This initiative developed the first reporting framework for the preparation of integrated reports. The organization was later rebranded into the International Integrated Reporting Council. This global alliance comprises investors, company representatives, and regulators, among others. The target of this initiative is to refine and disseminate the integrated reporting framework, which was published at the end of 2013 (Eccles et al., 2019). In addition, the IIRC wants to achieve superior information quality, which is accessible for "providers of financial capital" (International Integrated Reporting Council, 2013, p. 4). Proponents of this reporting method argue that it allows stakeholders to better understand the interconnections between the traditional reporting content and sustainability issues (Melloni, Caglio, & Perego, 2017).

To sum up, the integrated reporting framework is a frequently used guideline for integrated reporting and provides businesses with instructions on merging financial and sustainability reporting. Additionally, it also includes the Guiding Principles, which act as the basic concept for preparing and presenting the reporting content, and it also includes the Content Elements. The latter are questions, for example, concerning the business model and the governance structure to provide businesses with reporting guidelines. However, according to the International Integrated Reporting Council (2013), they should not serve as a rigid reporting structure since the content depends on the individual company situation. In this thesis, all annual reports, which contain information on the companies' sustainability issues, are defined as integrated reports. However, for the purpose of this study, it is not relevant whether these reports are in accordance with the integrated reporting framework.

4. Methodology

4.1. Introduction to Textual Analysis

In this paper, computer-based textual analysis of standalone and integrated reports is conducted using R. The reports are from companies that are a part of the STOXX Europe 600. R is a free programming language and environment with regards to the areas of graphics and statistical computing. In general, textual analysis can be defined as the "notion for parsing text for patterns" and is also known as natural language processing or computational linguistics (Loughran & Mcdonald, 2016, p. 1). This method comprises numerous techniques like sentiment analysis or the measurement of document similarity. Thereby, it allows researchers to extract information from unstructured data types, such as annual reports or other company statements (Liu, Wu, Yang, & Yu, 2020; Loughran & Mcdonald, 2016).

At the beginning of the 20th century, manual textual analysis was utilized. The motivation was, for example, to analyze works of Shakespeare or political speeches during the world wars (Loughran & Mcdonald, 2016). However, this manual method is time-consuming and prone to subjectivity. Hence, it is not recommended for the analysis of large-scale data samples (Yang, Dolar, & Mo, 2018). Computer-based textual analysis overcomes these problems since it is based on algorithms, which are not affected by subjective judgments (Mittelbach-Hoermanseder et al., 2019).

Moreover, in recent years, computing power has rapidly increased through technological innovation. In turn, this expanded the application opportunities of computer-based textual analysis. Technological progress, combined with the online availability of accounting- and finance-related documents, has led to the increasing use of this research method in the fields of accounting and finance (Loughran & Mcdonald, 2016). Thus, previous analyses were mainly concerned with, for instance, the degree of CSR disclosure in annual reports (Mittelbach-Hoermanseder et al., 2019), textual analysis of CSR reports in the US (Clarkson et al., 2020), and disclosure of sustainability information in annual reports (Hummel & Rötzel, 2019). All in all, textual analysis is a reliable and replicable form of qualitative analysis, which will be used in this paper to examine stand-alone and integrated sustainability reports, respectively.

4.2. Data Sample & Data Pre-Processing

The sample includes all firms that were part of the STOXX Europe 600 index at the beginning of June 2020. One has to consider changes in the index's composition, which came into effect on 22nd June 2020. These were caused by a regular review conducted by the index provider. Hence, companies such as Hugo Boss and EasyJet are still in the sample of observed companies, although they were excluded from the STOXX Europe 600 index through the last review (STOXX, 2020). In addition, as not all reports from 2019 have been published yet, the period between 2010 and 2018 is observed.

To find all relevant sustainability and integrated reports, the ASSET4 database, a division of Thomson Reuters, has been utilized. ASSET4 is a specialized ESG database, which contains around 900 indicators per year per company and provides information on the CSR performance of numerous businesses. This information is retrieved from publicly accessible sources and is predominantly related to the following four dimensions: economic, social, environmental, and corporate governance. Research analysts evaluate this information and create annual ESG scores for each firm (Clarkson et al., 2020).

A list indicating whether a company issued a sustainability report was retrieved from the ASSET4 database in the first step. The list also included an indicator of whether the reports are in accordance with the GRI framework. According to the database, 4,061 reports have been published between 2010 and 2018. Nevertheless, if all companies had reported for each year, 5,400 reports would be available (Appendix 1).

In the next step, the reports were retrieved as .pdf documents. Following the approach of Clarkson et al. (2020), the number of reports per year per was restricted to one observation per company. In some cases, for example, companies published an additional summary report of their sustainability report. In this case, the document with the highest number of pages was selected if more than one report per year was available. The reports were downloaded from the following sources: GRI database, company websites, and www.corporateregister.com. The latter one is the leading organization that provides CSR reports in its database (Dhaliwal, Li, Tsang, & Yang, 2011).

However, only 3,187 reports could be downloaded (Appendix 1). The primary issue was that many companies remove their sustainability reports after a certain time period. Furthermore, other reasons, which led to the exclusion, will be described subsequently. While some companies published only an executive summary instead of the full report (e.g., IHG's report in 2017), other companies published the report in a language other than English (e.g., Gecina's bilingual reports). Moreover, in some cases, false reports were published on the corporate website. Kesko, for instance, attached the sustainability report of 2012 instead of their report of 2010. Furthermore, some companies pursue a biennial reporting approach and, hence, published one report covering two years (e.g., Voestalpine). Since both years are covered, the ASSET4 database indicates that two reports are available. In turn, this led to a distortion of the initial sample size, which seems to be smaller than predicted. In the case of a company merger, the reports prior to the merger were excluded from the sample as, otherwise, the sustainability reports of either of the merging parties would have to be selected (e.g., EssilorLuxottica). Lastly, some companies published only a web version of their sustainability report, which could not be downloaded (e.g., Tele2 2013-2016).

During the gathering of the sustainability reports, it was noticed that there are some deviations from the ASSET4 database. Not all reports that were declared as GRI reports are indeed in accordance with the GRI framework. These errors were manually revised. However, these incorrectly declared reports account for only ∼1.2% of all sustainability reports.

Before the actual text analysis, one had to pre-process the integrated reports since they contain not only sustainability information but also other operational and financial information like balance sheets and income statements. To increase the comparability with the stand-alone reports, the sustainability part was manually extracted from the integrated reports. However, some reports could not be edited since the reports were protected (e.g., Coca Cola and Thyssenkrupp). Hence, these reports were excluded from the overall sample. This also applies to reports for which it was not possible to identify an extractable sustainability part because the individual sustainability components were spread across the entire integrated report. Afterwards, the sample comprised 2,619 reports (Appendix 1).

Nevertheless, the data sample still contained reports from

companies that did not report over the entire period or for which not all reports were available. Some companies have started their reporting activity in later years, for instance, due to the pressure of the EU Directive. However, the inclusion of reports from companies that have published only one or two reports might bias the analysis. Hence, to consider both aspects, the minimum number of reports per company was set at four. This limit guarantees that companies that have started reporting no later than 2015 can be included in the analysis. In the end, this resulted in a final sample of 2,431 reports (Appendix 1). Finally, the reports were imported into R. For this purpose, the package *readtext* was utilized.

In the next step, the data had to be pre-processed to accurately analyze numerous reports and determine the different textual characteristics. However, not all steps are relevant for the analysis of all observed variables. Thus, all steps will be described in the following part, and individual deviations for specific variables will be mentioned later. Moreover, the *quanteda* package was utilized. This package in R is predominantly determined for the quantitative analysis of textual data.

Firstly, the text was tokenized. During this process, the original text is broken down into a sequence of its individual components. These components are called tokens. There are different tokenization approaches. In this thesis, the text body was tokenized at non-letters. As a result, at every space or punctuation, a new token was created.

Secondly, specific characters were removed from the text body to improve the accuracy of the analysis. For this purpose, hyphens, punctuation, and symbols were filtered out. On top of that, in some cases, numbers were removed from the text.

Thirdly, all tokens were converted to lower cases. In turn, this ensures that identical words are identified as one word. Otherwise, words in different cases are not equivalent to each other and will be treated separately.

Fourthly, English stopwords were filtered out. These words are generic terms that have a low informational value and are irrelevant for the analysis. The package *quanteda* contains a list of 175 predefined stopwords. Examples for this word category are words like “that”, “would”, or “the”. Moreover, this reduces the number of tokens, which diminishes complexity and optimizes the processing time of the analysis.

Lastly, a stemming procedure was performed. This process is a type of morphological analysis, which aims to transform words into their root form, for instance, removing suffixes like “ed” and “ing” (Liew et al., 2014). Thereby, terms like “went” and “going” are converted into “go”. This procedure has a positive impact on precision since different versions of the same word are recognized as the same term. In addition, it further consolidates the dataset by reducing the number of tokens and, thus, positively affects processing time.

4.3. Observed Variables

As the data has been collected, imported, and pre-processed, the different observed variables can be analyzed. In the following sections, the variables will be defined, and the underlying methodology will be explained.

4.3.1. Length

The first observed variable is the length of disclosure. According to the current literature, the report length provides numerous insights. For example, the informational content might be derived from the length of CSR reports since longer reports can contain more information. This underlying argumentation has been utilized in various studies about financial and non-financial reports (e.g., Li, 2008; Muslu, Mutlu, Radhakrishnan, & Tsang, 2019). In contrast, it is also argued that the length might act as an indicator of the complexity of the companies’ CSR activities as these have to be more extensively described (Muslu et al., 2019). Moreover, Neu, Warsame, and Pedwell (1998) show that companies seek to influence the perception of stakeholders about the related CSR performance by disclosing environmental data. Hence, they argue that the length of a report also acts as a proxy for the management of stakeholders’ impressions.

In this thesis, three indicators are utilized to measure report length: number of characters, number of total words, and number of unique words. The first indicator does not require any pre-processing steps and counts all elements, including letters and numbers. The second indicator is the number of total words, which is the most common measurement type of report length in the current literature (e.g., Clarkson et al., 2020; Hummel & Rötzel, 2019; Muslu et al., 2019). Lastly, the number of unique words indicates how many different words are utilized in one report. The latter two indicators require all pre-processing steps except for the removal of stopwords. Following the approach of Hummel and Rötzel (2019), the logarithm of the measurement values is used to reduce skewness.

4.3.2. Readability

Readability is a parameter for the complexity of the language and determines “how easily the reader can grasp the content of a text” (Mittelbach-Hoermanseder et al., 2019, p. 27). Researchers often observed this variable in financial and non-financial studies. Managers might utilize the readability of corporate documents as an instrument to obfuscate a low performance since a difficult readability increases the burden of understanding. Thereby, it might diminish the resulting negative reaction of investors and analysts. Concerning CSR disclosure, the role of readability is especially pronounced due to the narrative and poorly regulated nature of CSR reports (Wang, Hsieh, & Sarkis, 2018).

Various indices have been adopted in previous studies for the measurement of readability (e.g., Clarkson et al., 2020; Hummel & Rötzel, 2019; Mittelbach-Hoermanseder et al., 2019; Muslu et al., 2019; Nazari et al., 2017; Wang et al., 2018). In this thesis, the Flesch-Kincaid, the Flesch Reading

Ease, and the Fog Index are utilized. According to Li (2008), these are the most reliable instruments. The calculation of these indicators is based on the average number of syllables per word and the average number of words per sentence. The Fog Index and the Flesch-Kincaid Grade Level estimate how many years of formal US education are required to understand the content of a text (Mittelbach-Hoermanseder et al., 2019; Wang et al., 2018). In addition, the Flesch Reading Ease score ranges from 0 (*Professional*) to 100 (5^{th} *grade*) (Flesch, n.d). Hence, there is an inverse relationship between both Flesch indices.

The calculation of the indices does not require any pre-processing steps. For the calculation, sentences with a minimum and maximum length of three and 75 tokens were considered, respectively. These limits ensure that incorrect sentence structures are excluded, and precision is optimized. Following the approach of Franco, Hope, Vyas, and Zhou (2015), the three indices were combined into one aggregate measure. This was conducted by computing the average of the indices' percentile ranks and dividing the result by 100. Thereby, higher values reflect a lower degree of readability.

4.3.3. Tone

By conducting a sentiment analysis, the tone of a report is determined. This variable indicates how positive or negative the document's language is, but it is not directly related to the content. Hence, it does not indicate, for example, whether good or bad news are published. However, even if the tone is not directly related to the content, sentences with a negative tone are likely to pertain negative information (Muslu et al., 2019). Due to the narrative nature of sustainability reports in contrast to financial reports, managers tend to utilize tone to communicate information, which is hard-to-quantify (Du & Yu, 2020). Referring to CSR disclosure, Muslu et al. (2019) argue that businesses with a more negative tone are more transparent since the negative tone relates to the publication of negative aspects. Hence, these companies are more willing to also inform the public about their negative impact on sustainability.

In order to conduct the sentiment analysis, a list of positive and negative words was utilized. This list was developed by Loughran and Mcdonald (2011) and is especially created for accounting research. All pre-processing steps of the text were conducted except for the removal of stopwords and the stemming process. Afterwards, all positive and negative words were counted. Finally, the tone was calculated by dividing the difference between positive and negative words by the total number of words (Figure 1). Thus, the value ranges between minus one and one.

$$\frac{Positive\ Words - Negative\ Words}{Total\ Words}$$

Figure 1: Sentiment Analysis – Formula

4.3.4. Topic-Specific CSR Disclosure

Many researchers investigated the degree of CSR content in company reports (Melloni et al., 2017; Muslu et al., 2019; Nazari et al., 2017). Some studies argue that CSR disclosure is utilized as a legitimizing tool leading to an increasing disclosure of sustainability information of companies with poor CSR performance (Cho & Patten, 2007). In contrast, Al-Tuwaijri, Christensen, and Hughes (2004) argue that companies with a superior CSR performance tend to disclose more CSR information. The underlying reason is that these companies want to convince market participants of their superior performance.

This thesis aims to assess the disclosure of specific CSR topics. In particular, the three dimensions measured are the three dimensions of the GRI Standards: economic, social, and environmental. Some researchers utilized the frequency of specific keywords as a measurement method of CSR disclosure. However, in this paper, the approach of Mittelbach-Hoermanseder et al. (2019) was applied. Thereby, one word window per dimension was created. Each word window contains twenty terms related to the specific dimension. Afterwards, the cosine similarities between the predefined word windows and the various reports were calculated. Cosine similarity is calculated as the inner product of two vectors: one vector refers to the topic-specific standard, and the other refers to a sustainability report. Thereby, the relative word frequencies were compared. This calculation resulted in the similarities between the reports' vocabulary and the topic-specific word windows (Mittelbach-Hoermanseder et al., 2019). The outcome ranges between zero and one. The latter result indicates that both documents have equal proportions, and the former result means that the documents do not share any similarities (Lang & Stice-Lawrence, 2015). Hence, a high value indicates that a report has a high similarity with the topic-specific vocabulary and that the company discloses a high degree of topic-specific information.

Every word window contains twenty words, including the respective search term and nineteen additional terms. However, "environment" is not included in the corresponding word window due to the term's ambiguity and is replaced by the term "ecology". The word windows are based on the ones defined by Mittelbach-Hoermanseder et al. (2019). Since Mittelbach-Hoermanseder et al. (2019) measured the five dimensions of the EU Directive, the word windows were consolidated and complemented by other terms. The additional terms were retrieved from the different standards. For example, the word-window of the economic dimension contains terms such as "corruption" (GRI-205), "monopoly" (GRI-206), and "tax" (GRI-207). The entire set of terms of the word windows and the corresponding GRI Standards are provided in Appendix 5.

4.3.5. Numeric Content

Numeric content in corporate statements allows the reader to gain insights into whether companies only talk about their activities or substantiate this with quantitative content and key performance indicators (KPIs). Researchers

argue that numerical information, in comparison to qualitative information, tends to be "more accurate, objective, comparable, and verifiable" (Hummel & Rötzel, 2019, p. 30). This information makes it easier for investors and analysts to understand the content of the reports and compare it with the ones of other companies. This is, for example, shown in the study of Huang, Nekrasov, and Teoh (2012). This study provides evidence that numbers in the title of earnings press releases cause a stronger reaction of investors.

In this paper, the indicator representing numeric content is the quantity of Arabic numbers per 1,000 words. Hence, it is measured as the ratio of the quantity of Arabic numbers to the total sum of words (Figure 2). However, it must be stressed that not the amount of digits is measured, but the amount of numbers. Thus, 100 is counted as one number and not as three digits. In addition, no pre-processing steps are necessary.

$$\left(\frac{Arabic\ Numbers}{Total\ Words}\right) * 1000$$

Figure 2: Numeric Content – Formula

4.3.6. Horizon Content

The degree of horizon-related information reflects the future orientation of corporate reports. Muslu et al. (2019) argue that sustainability reports, which elaborate on future trends, tend to be more informative for market participants. Referring to the research on financial reports, researchers provide evidence that forward-looking Management Discussion and Analysis (MD&A) statements are more informative and help market participants to consider the company's future performance in their analysis (Muslu, Radhakrishnan, Subramanyam, & Lim, 2015).

The approach of Muslu et al. (2019) was applied to calculate the degree of horizon content. The number of all horizon-related words was counted using the predefined list of Muslu et al. (2019). This list includes short- and long-term related phrases like "next period", "subsequent quarter", and "upcoming month" (Appendix 6). Lastly, the number of horizon words was divided by the total number of words and multiplied by 1,000 to receive the amount of horizon content per 1,000 words (Figure 3). Regarding the pre-processing steps, all of them were conducted except for the removal of hyphens, numbers, punctuation, and stopwords, and the stemming process. The underlying reason is that the list of Muslu et al. (2019) includes hyphens and numbers and is not stemmed.

$$\left(\frac{Horizon\ Words}{Total\ Words}\right) * 1000$$

Figure 3: Horizon Content – Formula

4.3.7. GRI Index

As described in the introductory part, every company that publishes a report following the GRI Standards has to release a GRI index. This document contains all topics mentioned by the company in its report. The total number of included topics acts as an indicator of the breadth of the report. Many researchers have already investigated the GRI index during their studies: Hummel and Schlick (2016) created a measurement scheme based on the GRI index to measure reporting quality, while Clarkson et al. (2008) developed a content index and incorporated components of the GRI index. For this study, the GRI index for each report is approximated to estimate the number of GRI items covered. One could also manually retrieve the reported items. However, this approach is relatively time-consuming. In addition, the approximation approach allows the estimation of a GRI index of companies that did not release a report in accordance with the GRI Standards.

In the first step, a dictionary for each GRI item was developed (Appendix 7; Appendix 8; Appendix 9). These act as search terms to identify whether a company has reported on the specific issues. Since generic terms such as "supplier" are mentioned in many buckets, for instance GRI-204 and GRI-308, the terms had to be very specific. In the next step, the only pre-processing steps were the removal of symbols and the transformation to lower tokens. Afterwards, the terms for each report were counted. As the single mention of a word does not mean that the company has reported on this item, thresholds were set. For items having more than five words in their dictionary, the limit was set at five. Therefore, the words had to occur a total of five times to exceed the threshold and confirm the item. For the remaining categories, a single mention is sufficient, as these are usually very specific. This becomes clear if one looks at the integrated report of Zalando in 2017. Even though they reported on "GRI-418 Customer Privacy", the report includes the terms of the corresponding dictionary ("customer privacy", "customer data") only four times. Finally, the total number of items confirmed was counted, ranging from zero to 34 and approximating the reporting breadth.

4.3.8. Target Orientation

Target setting is a fundamental element of management control systems (Malmi & Brown, 2008). These systems help companies break down their strategy into clear objectives for the different management levels and facilitate the execution of the company's strategy (Kaplan & Norton, 2008). Hence, target setting is a steering instrument and allows the management team to align the activities of the company with the desired organizational outcome (Malmi & Brown, 2008). Therefore, it is interesting to investigate whether companies formulate clear objectives in their reports. A precise formulation of sustainability objectives might indicate a higher commitment to a sustainability strategy.

For this paper, the degree of target orientation is measured by the number of target-related words per 1,000 words. Firstly, a list was created containing fifteen terms, which are

associated with target orientation. For this purpose, the first step was to search for synonyms for "target". As a result, terms such as "goal" and "objective" were found. In addition, for a test sample of ten reports (Appendix 10), the *keyword-in-context* function was used. This function allows searching for terms that are often mentioned in connection with the terms above. Furthermore, the 100 most frequent terms were searched for in the same test reports. All values that matched from a personal assessment were added to the list. In the last step, the set was supplemented by additional words that matched from personal experience resulting in the final dictionary (Appendix 11). Afterwards, the number of target-related words per report was counted and divided by the number of total words. To receive the number of target-related terms per 1,000 words, it was multiplied by 1,000 (Figure 4).

$$\left(\frac{Target\ Words}{Total\ Words}\right) * 1000$$

Figure 4: Target Orientation – Formula

5. Literature Review & Hypothesis Development

The following part delivers a comprehensive analysis of earlier results obtained from the existing literature. After a thorough evaluation of the different perspectives, the hypotheses concerning the relationship between the different variables and CSR performance will be formulated. The hypotheses relate to the textual features and the following additional factors: GRI framework, report type, and adopter type.

Referring to the length of disclosure, Clarkson et al. (2020) provide evidence that companies with superior CSR performance tend to disclose more comprehensive sustainability reports. The reason for this might be that well-performing companies can include more content since they pursue more CSR activities (Clarkson et al., 2020).

However, Li (2008) examined the relationship between the length of annual reports and financial performance and came to a different conclusion. The researcher mentions that longer reports lead to higher information-processing costs. Therefore, market participants need to invest more resources to understand the content of these reports accurately. Hence, managers might use the length as a strategic measure to decrease transparency and conceal information from market participants (Li, 2008). This strategy is also emphasized by Aureli et al. (2016), who state that companies can merely repeat information to increase the length of reports without providing additional and meaningful content. This insight underlines that longer reports can serve the purpose of disguising poor performance. Furthermore, this finding can be easily applied to CSR reporting. Companies could also increase the length of their sustainability reports to hide information about their poor CSR performance in an enormous amount of less relevant information. With regard to financial reporting, the US Securities and Exchange Commission (SEC) even recommends keeping sentences and documents as short and concise as possible to avoid this problem (SEC, 1998).

Current literature also often refers to the legitimacy theory (e.g., Clarkson et al., 2008; Clarkson, Overell, & Chapple, 2011; Patten, 2002). This theory states that the social legitimacy of businesses is monitored through a process of public policy. In case a company assumes that its social legitimacy is endangered, it is incited to approach the problem proactively. Thus, companies with a low CSR performance might utilize sustainability reporting as a legitimizing tool to positively affect the public perception about the company's performance. As a result, they report in detail on their sustainability issues to justify their legitimacy, which leads to longer reports. In addition, managers can employ this tool to inform the public about actual performance changes (Patten, 2002). Hence, this might lead to longer reports from companies with poor CSR performance.

Moreover, researchers also often refer to the voluntary disclosure theory (e.g., Al-Tuwaijri et al., 2004; Clarkson et al., 2008, 2011). In contrast to the legitimacy theory, it states that businesses with superior CSR performance have an incentive to disclose a high level of information. Thereby, these companies can highlight their performance and differentiate themselves from the competition. Competitors with poor performance cannot easily imitate this. This condition can alleviate the problem of adverse selection since it allows companies to reveal their actual performance level, which is not directly accessible to market participants. These companies might expect that market participants will react positively to this (Clarkson et al., 2008). In turn, this fact might lead to longer sustainability reports from well-performing companies. Hence, the voluntary disclosure theory suggests that CSR performance positively correlates with the length of sustainability reports.

The findings of the current literature about the relationship between length and CSR performance are inconclusive. Nonetheless, based on the insights gained through the voluntary disclosure theory, the hypothesis is formulated as follows:

> *Hypothesis 1: The length of sustainability reports is positively correlated with CSR performance.*

Next, the relationship between CSR performance and readability is examined. In general, businesses have an incentive to incorporate negative and positive sustainability information into their reports to preserve a positive company image. Otherwise, the disclosure of merely positive information would weaken the report's credibility and damage the company's reputation. Therefore, managers might utilize readability as an instrument to mitigate the response to negative information (Wang et al., 2018). This strategy would mean that reports of underperforming companies are less readable.

For this, Nazari et al. (2017) refer to the Efficient Market Hypothesis (EMH). This concept states that the market price incorporates all publicly available information. However, the researchers argue that the EMH is constrained by market participants' cognitive capabilities to retrieve all available information. Hence, they propose the Incomplete Revelation Hypothesis (IRH) of Bloomfield (2002). This theory states that market participants do not immediately react to complex information due to their cognitive limitations. Hence, the cognitive limitation of market participants allows managers to diminish the negative impact of bad news by worsening the readability since this makes it more difficult to accurately comprehend the information (Nazari et al., 2017; Wang et al., 2018).

Concerning financial reports, this obfuscation strategy can even inhibit the ability of small investors to process the reports since it requires too many resources (Miller, 2010). Thus, the usage of readability as a tool to obfuscate a poor performance was also observed in financial reports. Li (2008) found that poor readability is associated with low financial performance. In this way, companies try to hide their poor performance.

Referring to the voluntary disclosure theory, one could also assume that companies with superior CSR performance aim to make the report as comprehensible as possible. Hence, they tend to utilize plain language, which can be easily processed to underline their superior performance (Wang et al., 2018).

However, the study of Clarkson et al. (2020) came to the opposite conclusion. Their study revealed that reports from companies with good sustainability performance are less readable. The researchers argue that reports from these companies include "more sophisticated analyses", which are more difficult to understand (Clarkson et al., 2020, p. 21).

Nevertheless, the majority of current literature supports the hypothesis that reports of companies with poor performance are less readable. Since a high value of the variable indicates a low level of readability, the second hypothesis is the following:

> *Hypothesis 2: The readability of sustainability reports is negatively correlated with CSR performance.*

Referring to the literature on tone, Davis and Tama-Sweet (2012) investigated the role of this linguistic feature in earnings press releases. They found a positive correlation between the tone indicator and the subsequent ROA. In addition, a further study revealed that a more pessimistic tone in the MD&A section is associated with lower future ROA (Davis & Tama-Sweet, 2012). Hence, financial research revealed a positive correlation between tone and the subsequent financial performance.

Concerning sustainability performance, the study of Cho et al. (2010) indicates that firms aim to manage the perception of stakeholders by biasing the verbal tone in their sustainability reports. The reports from bad CSR performers reveal a high level of optimistic language compared to the opposite group. This finding shows that underperforming firms tend to focus on the good news while blurring the bad ones (Cho et al., 2010). This phenomenon is especially pronounced for sustainability reporting since it is still unregulated and difficult to verify (Du & Yu, 2020). Thereby, these findings support a negative correlation between both factors.

Nevertheless, the study of Clarkson et al. (2020) came to the opposite conclusion that the tone in reports from companies with poor CSR performance tends to be more negative. This result is not in line with the predicted greenwashing behavior. Greenwashing can be defined as "a discrepancy between words and deeds, which combines poor environmental performance and positive communication about the environmental performance" (Pizzetti, Gatti, & Seele, 2019, p. 2). Clarkson et al. (2020) argue that their finding, which contradicts the expected greenwashing behavior, might be caused by the attempt of managers to enhance credibility. Alternatively, they state that this might be a result of an unconscious process to fend off criticism. Nonetheless, they point out the need for further research (Clarkson et al., 2020). Moreover, Du and Yu (2020) found that improvements in tone indicate an improvement in subsequent sustainability performance. As managers employ positive and negative words to inform stakeholders about their expectations of future performance, an improvement in tone acts as an indicator for a higher future CSR performance. One condition for this assumption is that sustainability reports contain relevant information about future performance (Du & Yu, 2020).

The findings of the various studies are divergent. Nonetheless, the third hypothesis follows the concept of greenwashing, meaning that well-performing firms aim to shape stakeholders' perception by biasing the tone in their reports:

> *Hypothesis 3: The tone of sustainability reports is negatively correlated with CSR performance.*

The observed indicators *GRI Index* and *Topic-Specific CSR Disclosure* refer to the content of sustainability reports. For this reason, the literature review is performed for both categories together. Ingram and Frazier (1980) investigated the relation between environmental disclosure content and the underlying environmental performance. While content is measured by twenty categories like "Regulatory Compliance" or "Environmental Control", the environmental performance is measured by an index of the Council of Economic Priorities (CEP). The results imply that the content does not correlate with environmental performance (Ingram & Frazier, 1980). Other researchers came to the same conclusion (Wiseman, 1982). Nonetheless, it could be argued that the issue did not have the importance that it has today. This would explain diverging results.

Moreover, Patten (2002) identified several flaws in the previous studies; for example, no other control variables were considered, and the indicators for the environmental performance were not appropriate. Hence, his study considered these issues. The study revealed a negative correlation between the content of disclosure and environmental performance. This finding corresponds to the study of Cho and

Patten (2007), who concluded that firms utilize disclosure as a legitimizing tool. Clarkson et al. (2008) also refer to the legitimacy theory. This theory suggests that firms include more content in their sustainability reports to justify their poor performance and change the public's perception of their actual performance (Clarkson et al., 2008).

However, the study of Clarkson et al. (2008) is in line with the voluntary disclosure theory. As described previously, firms with superior CSR performance could incorporate more content to differentiate themselves from the competition. This proactive strategy allows well-performing companies to disclose their true performance, which, otherwise, is not directly accessible to market participants. These companies expect a positive market response and believe that this will lead to future benefits. Hence, this theory suggests a positive relation between CSR performance and the amount of content (Clarkson et al., 2008). The study of Al-Tuwaijri et al. (2004) revealed similar findings suggesting that firms with superior CSR performance tend to cover more topics in their sustainability reports.

To sum up, the various studies provide mixed results about the relationship between CSR performance and CSR disclosure. Moreover, one has to underline that many studies only observed the relationship between environmental disclosure and the associated performance. Nevertheless, the environmental dimension is only one element of sustainability and does not reflect the overall concept. In turn, the following hypotheses are formulated in line with the voluntary disclosure theory mentioned previously:

> *Hypothesis 4a: The number of GRI items per sustainability report is positively correlated with CSR performance.*

> *Hypothesis 4b: The degree of topic-specific CSR disclosure is positively correlated with CSR performance.*

The literature on the relationship between CSR performance and numerical content in sustainability reports is limited. Clarkson et al. (2011) examined the relation between the nature of disclosure and CSR performance. According to this study, the nature of disclosure can be characterized as "soft" or "hard". "Hard" disclosure is defined as information that is "objective and externally verifiable" (Clarkson et al., 2011, p. 2). This definition can also be applied to the concept of numerical content since, in comparison to narratives, it is more objective and easier to verify. In their study, a sample of 51 listed Australian companies is observed. While environmental performance is measured by the quantity of toxic releases weighted by annual sales, the degree of "hard" disclosure is measured by the ratio of "hard" disclosure items to the total number of disclosure items. The latter is based on an index developed by Clarkson et al. (2008). Although the researchers expected a positive relationship between "hard" disclosure items and environmental performance, the study revealed a negative relationship (Clarkson et al., 2011).

Moreover, one can also refer to the voluntary disclosure theory, as well as to the legitimacy theory, to formulate a hypothesis. Concerning the former one, this theory predicts that companies with superior CSR performance disclose CSR information as a unique selling proposition to differentiate themselves from underperforming companies. Hence, they might substantiate their reports with more numerical content, which is more credible, objective, and verifiable than narratives. This content cannot be imitated by firms with poor performance and underlines the outstanding performance. Therefore, this theory suggests a positive correlation between numerical content and CSR performance (Clarkson et al., 2011).

With reference to the legitimacy theory, underperforming companies publish CSR reports to maintain their legitimacy and positively affect the public's perception of the company. Therefore, this theory predicts that companies disclose less numeric content since this could harm the legitimacy. In turn, they prefer to publish a high degree of narratives because they aim to shape the public's perception and not communicate their actual performance (Clarkson et al., 2011).

Moreover, Hummel and Schlick (2016) examined the relationship between the quality of sustainability reports and the associated performance. High-quality disclosure is defined as the "disclosure of numerical data on a company-wide level that fulfill[s] or exceed[s] the minimum requirements derived from the GRI guidelines" (Hummel & Schlick, 2016, p. 460). In contrast, low-quality disclosure does not fulfill the requirements or provides any other information. In particular, the latter aspect demonstrates that low-quality disclosure tends to contain less numeric content when companies provide any other information than required. Moreover, the researchers argue that high-quality disclosure fulfills the following criteria: verifiability, comparability, and reliability. Numeric content can be easily compared among companies, is more reliable than narratives, and can be better verified. These characteristics also underline that numeric content can be defined as high-quality disclosure. The results of the study indicate that high-quality disclosure is positively related to CSR performance (Hummel & Schlick, 2016). Hence, one can assume that the relation between numeric content and CSR performance is also positive. In this paper, the hypothesis states as follows:

> *Hypothesis 5: The numeric content of sustainability reports is positively correlated with CSR performance.*

Current literature has not yet investigated the relationship between horizon content and sustainability performance. However, Muslu et al. (2019) examined the impact of CSR report narratives on the accuracy of analyst forecasts. Their results reveal that the quality of CSR reports, measured by a disclosure score, positively affects the analyst forecasts' precision. The researchers developed the associated disclosure score based on the following criteria: tone, readability, length, numerical content, and horizon content. Concerning the latter component, they argue that reports tend to

be more informative when they elaborate on the future outlook (Muslu et al., 2019). Muslu et al. (2015) support this assumption since they found that MD&A disclosures with a high degree of horizon content help market participants to forecast financial performance. Moreover, Hussainey and Walker (2009) examined the relation of forward-looking statements in annual reports and market participants' ability to predict future earnings. Their results also indicate that forward-looking disclosures improve the precision of analyst forecasts. Hence, it can be concluded that reports with a high degree of horizon content tend to be more informative.

This assumption can be utilized for the interpretation of the voluntary disclosure theory and the legitimacy theory. Regarding the former one, well-performing companies might aim to maximize the informativeness of their reports. The argumentation is based on the same train of thoughts as in the previous sections. Hence, reports from well-performing firms are likely to discuss the future CSR strategy and the associated activities. Thereby, these companies aim to differentiate themselves from underperforming firms. This theory suggests that horizon content is positively related to CSR performance.

Moreover, the implications of the legitimacy theory are inconclusive. On the one hand, one could argue that underperforming companies tend to keep the informativeness as low as possible since they do not want to disclose their misconduct. This behavior might be especially pronounced for underperforming companies, which do not want to change their sustainability strategy in the future. On the other hand, Asay, Libby, and Rennekamp (2018) studied the relation between firm performance and language choices in narrative publications. They argued that there are two arguments why these companies could incorporate more future-related content in their reports: Firstly, companies with a poor CSR performance might elaborate on future initiatives to distract from the past or use it as a measure of defense. Secondly, they might aim to affect the perception of the market positively. Therefore, these companies focus on what they will change during the next periods and how they will improve (Asay et al., 2018). This behavior would tend to apply to companies that want to improve their CSR performance in the future. Consequently, this theory does not propose a clear relationship between the current CSR performance and horizon content. However, since there is more support for a positive relationship, the hypothesis is as follows:

> *Hypothesis 6: The horizon content of sustainability reports is positively correlated with CSR performance.*

The amount of literature elaborating on the relationship between CSR performance and target orientation of sustainability reports is similarly limited. Targets are an essential part of management control in most organizations and act as a decision-making tool (Arnold & Artz, 2015). To develop a hypothesis about the relationship, one can refer to the goal-setting theory developed by Latham and Locke (1979). The researchers argue that challenging but attainable objectives, which are clear and specific, positively impact performance. This effect results from four mechanisms. Firstly, objectives direct the behavior and effort of employees towards measures, which serve the purpose of fulfilling the objective. Secondly, objectives motivate employees through the so-called "energizing function" (Locke & Latham, 2002, p. 706). Thirdly, objectives, in particular difficult ones, improve the persistence of employees and, fourthly, they indirectly affect employee behavior by generating knowledge and arousal (Locke & Latham, 2002). Thus, this theory suggests that the formulation of sustainability goals has a positive impact on performance. However, the mere number of words does not measure the content or the formulation of objectives since the theory is based on specific and attainable objectives. Moreover, this argumentation refers to a causal effect from the report design on CSR performance, while the opposite effect is observed.

As in the previous paragraphs, one can refer to the voluntary disclosure theory and the legitimacy theory. The first theory predicts a positive correlation between CSR performance and the degree of target orientation. Well-performing companies might tend to formulate clear objectives to show the market that they have met their targets in the following period. This action would enable them to demonstrate their superior CSR performance and set themselves apart from the competition. In contrast, companies with poor performance would formulate fewer or no targets at all, as they are unlikely to meet them.

From the legitimacy theory perspective, one could argue that underperforming companies might set targets to convince market participants of their legitimacy. According to the arguments provided previously, they might even set themselves more targets to distract from the current situation. As a consequence, the legitimacy theory does not provide a consistent prediction in line with the goal-setting theory or the voluntary disclosure theory. Nonetheless, the present hypothesis is based on the latter perspectives:

> *Hypothesis 7: The target orientation of sustainability reports is positively correlated with CSR performance.*

In addition to the textual characteristics, the effect of other variables is observed. An additional factor is whether the company is an early or late adopter of sustainability reporting. Early adopters are companies that pursued sustainability reporting before the announcement of the EU Directive in 2014. In contrast, late adopters are defined as companies that started reporting on their sustainability issues after the announcement until the regulation came into force. In the current literature on IFRS adoption, there is also a third group, the so-called resisters, which are businesses that started reporting after the application of the IFRS became mandatory (Christensen, Lee, Walker, & Zeng, 2015). In terms of the sustainability reporting environment, this would include companies that started reporting in 2017. At this

point in time, the first mandatory reports were published covering the financial year 2017-2018. However, this group is excluded from the current sample since the minimum number of reports per company is four. The resister group has only published two reports during the observation period for the years 2017 and 2018.

Bhimani, Silvola, and Sivabalan (2016) conducted interviews and surveys to examine the relation between the two reporter types and CSR embeddedness. The latter is defined as "alignment between the contents reported and their actual manifestation" (Bhimani et al., 2016, p. 82). Thereby, it acts as an indicator to what extent the sustainability reports' content is integrated into the overall strategy. Even if it is not directly related to CSR performance, embeddedness might be positively correlated with CSR performance when companies do not only talk about their CSR strategy but, in turn, implement it. The researchers argue that early adopters' motivation is related to their genuine ambition for CSR activities and their intention to talk about the underlying performance. In contrast, late adopters are only motivated to keep up with early adopters and merely improve their reputation. Hence, sustainability practices are more integrated into the strategy of early adopters leading to a higher sustainability performance (Bhimani et al., 2016).

This argumentation is consistent with the theory of voluntary disclosure. Firms with superior CSR performance aim to inform the public about their positive impact. As a result, they started reporting on sustainability issues without regulatory pressure. In contrast, companies with poor CSR performance do not want to share this information with the public and only start reporting in response to regulations. Without this pressure, they would probably hesitate to adopt the practice of sustainability reporting. Hence, this supports the hypothesis that early adopters obtain a higher CSR performance than late adopters.

Nonetheless, given the legitimacy theory, it could be argued that companies with poor CSR results started reporting even before the announcement of the directive. Since these companies are striving to positively shape the perception of their stakeholders and convince them of their legitimacy, they might have an incentive to start reporting without regulatory pressure. In contrast, high-performing companies do not face external pressure to justify their legitimacy and only start reporting in response to upcoming regulations. Thus, this supports the hypothesis that early adopters tend to be companies with poor CSR performance to justify their legitimacy. Nonetheless, since the voluntary disclosure theory is supported by the study of Bhimani et al. (2016), the hypothesis is as follows:

> *Hypothesis 8: Companies with a higher CSR performance tend to be early adopters of sustainability reporting.*

This paper also examines whether there is a relationship between reporting under the GRI framework and CSR performance. Bernard, Abdelgadir, and Belkhir (2015) conducted a sector-specific analysis to examine the relationship between the two factors. They utilized CO_2 emissions as a measure of sustainability performance in the period between 2007 and 2011. In their study, they found no performance differences between GRI and non-GRI reporting firms. Therefore, this study does not identify any significant effect. However, the study does not include any control variables, and it is questionable whether CO_2 emissions are a representative instrument for measuring sustainability performance.

Referring to the voluntary disclosure theory, one could assume that high-performing companies tend to commit themselves to the GRI framework to help stakeholders better understand the company's CSR activities. Moreover, underperforming companies probably avoid using the GRI guidelines, as they can be better compared to high-performing companies. Thus, companies with superior CSR performance can distinguish themselves even better from the competition through reporting under the GRI framework. Lastly, well-performing companies make their intentions even more apparent to stakeholders by committing to this framework since the preparation of these reports requires effort and resources.

With reference to the legitimacy theory, one could also argue that companies report under the GRI framework to compensate for their poor sustainability performance. In this way, they might try to signal their stakeholders that they want to work on themselves. These firms expect that a commitment to one of the leading reporting guidelines would compensate for the poor performance and has a positive effect on the stakeholders' attitude towards the company. Even if they cannot deliver sufficient CSR results, they justify their legitimacy by producing high-quality reports in line with the GRI framework (Bernard et al., 2015).

However, the literature on this relationship is limited, and there is little empirical evidence provided by studies. For this reason, it is only hypothesized that both factors are correlated, but not the direction:

> *Hypothesis 9: The CSR performance affects whether companies prepare their sustainability reports in accordance with the GRI framework.*

The last factor observed concerns whether companies publish a separate report or integrate the sustainability section into their annual report. However, no empirical study dealing with the differences among stand-alone or integrated reports could be identified. Therefore, the underlying argumentation is based on the voluntary disclosure theory and the legitimacy theory, respectively.

With reference to the former one, the line of argumentation is not clear. On the one hand, one could argue that high-performing companies publish a separate sustainability report highlighting their superior CSR performance. One report, which solely focuses on the company's impact on sustainability, underlines the company's efforts and raises stakeholders' awareness. Moreover, other companies might aim to hide their poor performance within their annual reports. Thereby, they can distract from their sustainability failure by focusing on financial performance. On the other

hand, one could argue that well-performing companies publish an integrated report highlighting the linkage between their CSR strategy and the operational business. These companies might want stakeholders to fully understand the company's impact and, hence, want to present financial and sustainability performance together.

Concerning the legitimacy theory, the argumentation builds upon the discussion on the GRI framework, previously. Companies with poor CSR performance might prepare a stand-alone report to protect their legitimacy. Even if they cannot deliver satisfactory results, they show their stakeholders that they are aware of their problems and wish to improve. For this reason, they prepare a stand-alone report underlining the company's efforts. However, one could also argue that these companies publish an integrated report. In this report, they can show the connection between their business model and their sustainability performance. Thereby, they might justify their performance with the firm's business model and claim that the poor performance, for instance, is due to the financial result or the operational complexity. By showing the linkage between business model and sustainability performance, they might try to defend their legitimacy.

To sum up, both theories predict that the CSR performance has a particular impact on the reporting method. However, the direction of the effect is unclear. For this reason, the hypothesis states as follows:

> *Hypothesis 10: The CSR performance affects whether companies publish integrated or stand-alone reports.*

6. Descriptive Statistics

Before the regression analysis is conducted, the descriptive statistics are presented subsequently. In addition to the results of the total dataset, the results for the following groupings will be compared: integrated vs. stand-alone reports and reports from early adopters vs. reports from late adopters. Besides, the changes over time will be examined. Referring to the length indicators, the absolute values and not the logarithmic values are presented since this facilitates the corresponding interpretation. Mann-Whitney-U tests are conducted to test whether the differences between reports published by early and later adopters, as well as the differences between integrated and stand-alone reports, are significant. This test is a non-parametric test for two independent samples, which checks whether they have an equal distribution. It is also often called the Wilcoxon-Mann-Whitney test.

6.1. Overview

The minimum, maximum, and mean values of all variables for the total dataset are listed in Table 1. Looking at the length indicators, one can detect enormous differences. While some companies report in great detail, other companies spend only a few hundred words on their sustainability reports. In general, an average sustainability report comprises around 27,000 words. Furthermore, the tone variable predicts that the reports are generally more positive than negative but almost neutral. This can be compared to annual reports, which usually have a lower value (Mittelbach-Hoermanseder et al., 2019). However, the dataset mainly comprises stand-alone sustainability reports (Appendix 2). Since annual reports are legal documents, they require a more neutral language, which might cause the difference between the two studies (Mittelbach-Hoermanseder et al., 2019).

The readability index cannot directly be interpreted since it represents the average of percentile ranks. The corresponding values cover the full range from zero to one. Looking at the individual readability indices, the Fog Index, as well as the Flesch-Kincaid Index, are above the normal range. Li (2008) mentions that the range of the Fog Index is from 8 (*childish*) to 18 (*difficult*), while all texts with even higher values are unreadable. For this reason, the average Fog score of this sample (21) seems to be out of range. This finding can also be observed for the Flesch-Kincaid Index. Since both scores represent the required years of formal education, it also underlines that the mean values might be too high. However, referring to a study of Caglio, Melloni, and Perego (2020), their results also reveal a high average Fog score of approximately 23. Hence, the results indicate that a high academic level is required to understand the content of sustainability reports. This may be based on the use of several technical terms related to the topic of sustainability, which might not refer to individual's common way of speaking. Concerning the Flesch Reading Ease, the results can be interpreted. The mean value of this readability index corresponds to a required educational level of a college graduate (Flesch, n.d). This insight is in line with the other two readability indices, which indicate that a high educational level is required.

The three indicators for measuring the disclosure of specific CSR topics reveal that social issues are the most prevalent topic in the various reports of the dataset. This is in line with the findings of Mittelbach-Hoermanseder et al. (2019), who also utilized the cosine similarity. The researchers observed that social is one of the predominant issues in annual reports. However, their study revealed that the cosine similarities of the social and environmental topics are lower in absolute values. Since their research solely investigated the content of annual reports and not sustainability reports, the diverging results may be caused by sample differences. The degree of CSR topics in annual reports is probably lower because the main focus of these reports is on financial and operational issues.

Referring to the GRI index, some companies cover all GRI items, while others do not even cover a single item. Since the sample includes reports, which are not in accordance with the GRI framework, this might explain the finding. However, on average, companies report on approximately 50% of the GRI items.

Table 1: Descriptive Statistics – Overview

Variable	Range	Minimum	Mean	Maximum
#Characters	0 – ∞	4,081	319,472	2,534,315
#Total Words	0 – ∞	417	27,065	205,362
#Unique Words	0 – ∞	169	2,247	6,859
Numeric Content	0 – 1,000	12.0	87.1	461.4
Horizon Content	0 – 1,000	0.0	1.0	5.9
Target Orientation	0 – 1,000	0.0	6.4	28.4
Tone	(1) – 1	(0.018)	0.006	0.035
Readability Index	0 – 1	0.0	0.5	1.0
Flesch Reading Ease	0 – 100	4	23	73
Fog	yrs. of edu.	15	21	26
Flesch-Kincaid	yrs. of edu.	12	17	22
Economic	0 – 1	0.000	0.066	0.178
Environment	0 – 1	0.000	0.146	0.385
Social	0 – 1	0.000	0.174	0.316
GRI Index	0 – 34	0	16	34

6.2. Development over Time

In recent years, it has been observed that more and more companies have begun to report on their sustainability responsibilities (KPMG, 2017). For this reason, it is of great interest to investigate how sustainability reports and their contents have developed over time. In particular, events such as the announcement of the EU Directive in 2014 and its subsequent entry into force in 2017 could have had an impact. Table 2 shows the mean values for all textual characteristics between 2010 and 2018.

Referring to the three length indicators, one can observe an increase in the average length after the EU Directive announcement. This effect is most pronounced for the number of characters and the total number of words. The corresponding plots can be seen in Appendix 12 to Appendix 14. Hummel and Rötzel (2019) investigated the impact of the introduction of the Companies Act 2006 Regulations 2013 in the UK on annual reports. This regulation demands the disclosure of various CSR information relating to issues like human rights and gender distribution. Their study also reveals that the implemented law led to an increase in the report length. The researchers argue that this can be explained by additional disclosure required by the regulation (Hummel & Rötzel, 2019). Hence, the increase in report length could be caused by the announcement of the EU Directive since businesses have started to report on additional topics not covered previously. This is further illustrated by the fact that the directive mandates the disclosure of the following four issues: environmental protection, human rights, anti-corruption, and social responsibility (European Union, 2014). However, one has to consider that the regulation was only announced but not implemented at that point in time. Looking at the literature on the mandatory IFRS adoption, the finding of this study is consistent with the study of Lang and Stice-Lawrence (2015). This study found that compulsory IFRS adoption led to an increase in the amount of disclosure.

In contrast, Chauvey, Giordano-Spring, Cho, and Patten (2015) argue that the increase in report length could be caused by more companies publishing stand-alone reports instead of integrated reports. However, in this study, the proportion of integrated reports increases over time, and these report types tend to be shorter (Appendix 24 – Appendix 27). Therefore, the proportion of integrated reports cannot be the underlying reason for this phenomenon. Moreover, the share of reports in accordance with the GRI framework is also almost constant, being no decisive factor, as well (Appendix 28). Even the emergence of late adopters due to the announcement in 2014 cannot be regarded as an underlying reason for the increasing length of reports, since their reports appear to be shorter than the reports from early adopters (Appendix 29 – Appendix 32). In summary, it can be assumed that the requirements of the EU Directive have led to an increase in the length of sustainability reports.

Regarding the horizon content, there is little movement over the observed period, and no specific trend can be identified (Appendix 16). The marginal increase in numeric content, shown in Appendix 15, may be explained by public demands for more performance indicators rather than narrative descriptions of the implemented policies. This is based on the fact that, in comparison to narrative descriptions, numeric content allows stakeholders to better assess the companies' CSR performance (Bhimani et al., 2016). Furthermore, Hummel and Rötzel (2019) provide evidence that the introduction of a similar regulation in the UK has also led to an increase in numeric content. This might be caused by the requirements of the directive.

On top of that, following the announcement of the EU Directive, the degree of words related to target orientation has increased (Appendix 17). As already mentioned in chapter five, the literature on this subject is relatively limited. However, the EU Directive mandates that reports must include tar-

gets for the implemented diversity policies (European Union, 2014). This regulation will inevitably have led to companies reporting more about the corresponding objectives. In addition, it could also have triggered that these companies started reporting on the goals of other CSR policies. Since legal regulations required them to set targets for their diversity policies, they might also have set targets for other policies. In turn, this might lead to an increase in target orientation.

Besides, the changes in tone are marginal and inconsistent (Appendix 18). This finding is consistent with the study of Hummel and Rötzel (2019), who similarly found no change in tone. Furthermore, it can be seen that readability has become increasingly difficult since 2014 (Appendix 19). In contrast, Hummel and Rötzel (2019) observed that the UK's regulatory intervention led to an improvement in readability. One reason for the observed negative impact in this study could be that the EU Directive does not provide clear reporting guidelines but leads to more disclosure of technical topics. In turn, this could negatively affect readability (Lang & Stice-Lawrence, 2015). However, as the requirements of the EU Directive are not very technical, it is questionable whether it indeed increased the complexity. Nonetheless, the EU Directive does not require reports to be based on particular frameworks but requires the disclosure of additional content. The combination of these aspects could be detrimental to readability if companies need to include more content in their reports without having appropriate guidelines for creating such a corporate document (European Union, 2014). Another reason could be the emergence of late adopters in response to the announcement. As these companies have not previously reported on their sustainability issues, their incentives for preparing a high-quality report could be relatively low. Another reason might be that late adopters are inexperienced in preparing such a report, thereby, harming readability.

With reference to the thematic disclosure indicators, it can be noted that the degree of economic content has continuously risen since the beginning of the observation period (Appendix 20). In contrast, environmental content has increased from 2014 onwards (Appendix 21), while the level of social content varies over the period and does not give a clear direction (Appendix 22). The findings are partially consistent with the results of Mittelbach-Hoermanseder et al. (2019), who found annual increases for the different topic-specific CSR disclosures.

Concerning the social dimension, one could argue that the EU Directive requires that companies report on social matters (European Union, 2014). Hence, this should have led to an increase in social content. Nonetheless, the average cosine similarity of the social dimension is already high in 2010, exceeding the economic and environmental values for the entire period (Table 2). Therefore, it is not surprising that the social value has not risen further since it was already a prevalent topic. Another reason, which could explain this insight, is the "comply or explain" clause (Ioannou & Serafeim, 2017, p. 3). The EU Directive allows businesses to justify why they do not pursue specific policies (European Union, 2014). Thereby, companies can avoid implementing such policies as well as the corresponding reporting, if they explain the underlying reasons. Hence, the announcement and the introduction of the EU Directive did not necessarily lead to a higher level of disclosure (Ioannou & Serafeim, 2017).

Furthermore, it is interesting that the mean value of the environmental dimension experiences an increase after the announcement of the EU Directive. This increase could be caused by the mandate to report on environmental matters like "the use of renewable [...] energy, greenhouse gas emissions, water use and air pollution" (European Union, 2014, p. 2). A similar insight was generated by the study of Hummel and Rötzel (2019), who observed an increase of information on required topics after a similar regulation became effective in the UK. Due to the fear of being caught to be non-compliant or due to perceived benefits of being compliant, companies might include more CSR information in their reports as a reaction to the EU Directive (Ioannou & Serafeim, 2017).

Concerning the constant increase of economic disclosure from 2010 onwards, no literature explains the effect before 2014. The economic dimension is related to topics like bribery, taxes, and minimum wages. On top of that, it can be observed that there is a particular trend towards integrated reporting since the share of integrated reports has continuously risen from 2010 onwards (Appendix 24). Since these report types combine business topics with sustainability issues, the CSR sections might include more content related to the economic dimension. Hence, the trend towards integrated reporting could be an underlying reason for the increase in economic content. Nonetheless, the subsequent chapter reveals that integrated reports contain less economic content than stand-alone sustainability reports. Since there is no specific explanation for this observation, this might be caused by the recognized trends towards sustainability reporting. Companies may have expanded their reporting scope by including other aspects than only environmental and social information.

The number of GRI items has increased after the directive entered into force in 2017 (Appendix 23). In the first step, the number of reports complying with the GRI framework is observed. Even if the GRI index is approximated for non-compliant companies, companies actively applying this framework might have a tendency to cover more GRI topics. However, the share of reports under the GRI framework remained almost constant during the observed period (Appendix 28). Therefore, this effect is not caused by an increasing number of companies reporting under the GRI framework. Another reason might be that companies have increased the content of their reports regardless of the framework following previous arguments provided on topic-specific CSR disclosure.

To sum up, it can be said that even if the EU Directive did not come into force until 2017, the mere announcement of the regulation was accompanied by significant changes to the textual characteristics of sustainability reports. Thus, the present study provides valuable insights on the development

of the key components of sustainability reports.

6.3. Integrated vs. Stand-Alone Reports

Another revealing observation is the comparison of integrated reports and stand-alone reports. Similar to the findings of Mittelbach-Hoermanseder et al. (2019), the sample indicates a trend towards integrated reporting since the proportion of integrated reports has increased over time, while the absolute number of reports also has risen (Appendix 3, Appendix 24). This trend is also consistent with the findings of Lukomnik, Kwon, and Welsh (2018), who observed that the number of reports in accordance with the integrated reporting framework doubled between 2013 and 2018. However, it should be noted that even if researchers have already conducted textual analyses on stand-alone and integrated sustainability reports, no study examined the differences between the two types of reports (e.g., Caglio et al., 2020; Mittelbach-Hoermanseder et al., 2019; Nazari et al., 2017).

Concerning the length indicators, stand-alone reports are generally longer than integrated reports (Table 3). This difference is significant at the 1% significance level for all three variables. In fact, they tend to be even twice as long if considering the number of characters and words only. Caglio et al. (2020) conducted a textual analysis on integrated reports in South Africa, and their results revealed that integrated reports include, on average, around 31,000 words. These integrated reports are twice as long as the integrated reports from this sample, which obtain, on average, approximately 15,000 words. Nonetheless, the researchers analyzed the overall report and extracted not only the sustainability part. Hence, the underlying methodological differences may explain the variations in the results.

One can refer to the audience of the different reports to explain the variance in length between the two reporting types. Yusof (2018) argues that the main target group of sustainability reports are stakeholders, while shareholders are the main target group of integrated reports. Furthermore, one could assume that companies tailor their reports to the needs of their audience. The study of Cohen, Holder-Webb, and Zamora (2015), based on a survey of more than 200 professional investors, revealed that investors prefer concise non-financial information. As a result, companies might keep their integrated reports short since the key audience are investors, who prefer brief rather than extensive reports. The integrated reporting framework also encourages this format by stating that “an integrated report should be concise” (International Integrated Reporting Council, 2013, p. 34). Thus, businesses that utilize this framework to produce their integrated report might follow this guiding principle and keep their reports short and concise. Another potential explanation is introduced by Yusof (2018), who investigated the change in sustainability disclosure when companies move from publishing stand-alone reports to publishing integrated reports. The study shows that after the integration, sustainability sections also tend to be shorter. The researcher argues that more information has to be included in a report and, hence, companies have to short financial as well as non-financial information to combine both parts (Yusof, 2018). This trade-off could also be a reason for this phenomenon.

Concerning horizon content and target orientation, there are only minor differences between the two reporting types. While the difference for horizon content is not statistically significant, the difference concerning target orientation is significant at 1%. However, it might be challenging to explain this phenomenon since stand-alone reports contain, on average, 0.3 target-related words more per 1,000 words. In addition, the tone of both report types is similar, and the marginal difference is not statistically significant.

Moreover, stand-alone reports tend to contain a higher degree of numeric content, which is significant at 1% (Table 3). Current literature does not provide any explanations for this observation. Nonetheless, it may be caused by the data extraction methodology. Since only the CSR sections from integrated reports were extracted, the tables might not be considered during the process. This is the case when the corresponding tables are located in subsequent chapters, such as the Appendix. Thus, the potential exclusion of tables from integrated reports might explain the difference in numeric content.

In terms of readability, integrated reports, on average, seem to be less readable than stand-alone sustainability reports (Table 3). This applies to the readability index as well as to the individual components. All of these differences are statistically significant at 1%. This finding is inconsistent with the study of Lueg, Lueg, Andersen, and Dancianu (2016), who found out that preparing an integrated report allows companies to use less technical language. In turn, this should lead to more readable reports. However, the current literature does not provide any reasons for the finding of this study. One possible explanation may be based on the combination of financial as well as sustainability issues. On the one hand, combining both parts leads to a certain level of complexity and, on the other hand, financial reports may require the use of more technical terms. Hence, these reasons might lead to a decrease in readability. Finally, this finding could also be potentially explained by the GRI framework. In Appendix 2, it can be seen that around 80% of the stand-alone reports comply with the GRI guidelines, while only about 47% of the integrated reports comply with the GRI framework. Therefore, integrated reports might be less readable due to the lack of guidance.

Concerning topic-specific disclosure, stand-alone reports are characterized by marginally higher economic content (Table 3). Besides, these reports also contain, on average, more content related to the environmental dimension. Both differences are statistically significant at 1%. This observation is consistent with the findings of Yusof (2018) and follows the argumentation that companies have to balance financial and sustainability issues in an integrated report. In turn, this can lead to a lower proportion of economic and environmental content. This insight could also be potentially explained by the study of Marx and Mohammadali-Haji (2014). The researchers examined integrated reports in South Africa and

Table 2: Descriptive Statistics – Development over Time

Variable	2010	2011	2012	2013	2014	2015	2016	2017	2018
#Characters (in k)	293	303	299	309	325	313	324	341	355
#Total Words (in k)	26	27	26	26	27	26	27	28	29
#Unique Words (in k)	2.3	2.3	2.2	2.2	2.2	2.2	2.2	2.3	2.3
Numeric Content	83.4	87.7	87.2	85.5	87.6	86.0	87.3	87.8	90.0
Horizon Content	1.0	1.0	1.0	1.1	1.0	1.0	1.0	1.0	1.0
Target Orientation	6.2	6.2	6.3	6.1	6.2	6.4	6.6	6.7	7.0
Tone	0.006	0.006	0.006	0.005	0.006	0.006	0.006	0.005	0.005
Readability Index	0.48	0.48	0.48	0.49	0.50	0.50	0.52	0.54	0.54
Economic	0.059	0.061	0.061	0.063	0.065	0.066	0.068	0.071	0.073
Environment	0.136	0.140	0.141	0.141	0.143	0.148	0.150	0.152	0.157
Social	0.175	0.172	0.174	0.172	0.174	0.174	0.173	0.176	0.173
GRI Index	15.4	16.0	15.7	15.7	15.7	15.6	15.8	17.2	17.4

found out that some businesses merely declared their annual reports as integrated reports without changing the reports' content. The underlying motivation is to pretend that they are on the cutting edge of sustainability reporting. As a result, this might lead to a low disclosure level of CSR topics in integrated reports (Marx & Mohammadali-Haji, 2014).

However, the results show that integrated reports tend to include more information on the social dimension than stand-alone sustainability reports. This difference is also significant at 1%. This contradicts the findings of Yusof (2018), who states that companies disclose approximately 70% less social information in their integrated reports. Nonetheless, his study focuses on reports from companies that have switched from stand-alone to integrated reports, while this study covers all report types (Yusof, 2018). Consequently, there is no empirical or theoretical finding explaining this observation.

Lastly, stand-alone reports cover around 50% more GRI topics than integrated reports (Table 3). Similar to the other observations, this difference is also statistically significant at the 1% level. As described above, most of the stand-alone reports follow the GRI guidelines. Companies that prepare their sustainability reports without relying on a specific framework might choose the focal points individually. In contrast, GRI-compliant companies probably focus on particular GRI topics. In turn, this might lead to the disclosure of more GRI topics and explains the differences in the reporting scope.

In summary, various textual differences between integrated and stand-alone reports can be detected. The main findings are related to the differences in length, readability, and topic-specific content.

6.4. Early Adopter vs. Late Adopter

The announcement of the EU Directive in 2014 led to the emergence of late adopters. As described above, these companies anticipated that the regulation would come into force in the following years and started reporting between 2014 and 2016 in response to the announcement of the EU Directive. All reports from companies that published fewer than four reports were excluded from the sample. Thus, in this study, the group of late adopters compromises all companies that started reporting for the financial years 2014 and 2015. The percentage of reports published by early and late adopters over time can be found in Appendix 29.

Concerning the average length of reports, it can be seen that the reports from early adopters tend to be more as twice as long as those from late adopters. This observation is statistically significant at the 1% level for all three variables (Table 4). It is also in line with the study of Stent, Bradbury, and Hooks (2013), who observed the impact of the transition to mandatory IFRS reporting in New Zealand. They also found that early adopters utilized about twice as much space for their annual reports compared to late adopters.

To explain this insight, one can refer to the motivational factors for voluntary reporting. The disclosure of CSR information can lead to various benefits such as lower cost of capital. In contrast, it is also associated with different costs, for instance, preparation and proprietary costs. If the benefits exceed the corresponding costs, firms are encouraged to voluntarily disclose CSR information (Ioannou & Serafeim, 2017; Leuz, 2010). Consequently, one can assume that the costs exceed the benefits for late adopters and, thus, they have not reported until the regulation was announced. This explanation is in line with the study of Christensen et al. (2015) about the mandatory adoption of the IFRS. The researchers argue that early adopters have strong financial reporting incentives, while late adopters adopt the framework in anticipation of the transition to compulsory reporting (Christensen et al., 2015). For this reason, one can assume that voluntary reporters obtain more significant reporting incentives to disclose CSR information. Since the perceived net benefit of sustainability reporting is minimal or even negative, late adopters might tend to keep their reports as short as possible. In this way, they could prepare for the incoming regulation and fulfill the upcoming minimum requirements. However, they keep their reports short to avoid the occurrence of additional costs.

Table 3: Descriptive Statistics – Integrated & Stand-Alone

Variable	Integrated	Stand-Alone	Sig. Level
#Characters	168,364	376,727	1%
#Total Words	15,106	31,596	1%
#Unique Words	1,581	2,500	1%
Numeric Content	80.7	89.5	1%
Horizon Content	1.0	1.0	n.s.
Target Orientation	6.2	6.5	1%
Tone	0.006	0.006	n.s.
Readability Index	0.60	0.47	1%
Flesch Reading Ease	22	24	1%
Fog	18	17	1%
Flesch-Kincaid	22	21	1%
Economic	0.061	0.067	1%
Environment	0.135	0.150	1%
Social	0.182	0.170	1%
GRI Index	11.5	17.8	1%

Mann-Whitney-U-Test (1%, 5%, 10%, or not significant (n.s.))

Moreover, it can be seen that early adopters, on average, incorporate more numbers into their sustainability reports (Table 4). This difference is significant at 5%. To explain this finding, one can refer to the companies' compliance with the GRI Standards. In Appendix 33, it can be seen that less than 50% of the reports published by late reporters are in accordance with the GRI framework, while about 75% of the reports published by early reporters comply with these guidelines. Bhimani et al. (2016) provide evidence that early adopters pursue a differentiation strategy to distinguish themselves from their competitors. This strategy could lead to a tendency for them to report under the GRI framework to set themselves further apart. Looking at the individual GRI disclosure items' requirements, these often demand the publication of numerical data. The GRI-303 standard, for example, requires the disclosure of the amount of water extracted as well as the amount of water recycled (Global Reporting Initiative, 2016d). Moreover, plotting the data in a diagram with the number of GRI items on the x-axis and numeric content on the y-axis, the trend line shows a tendency of the number of reported GRI positions to positively correlate with the degree of numeric content (Appendix 34). This correlation underlines that reports with many covered GRI topics include, on average, more numerical data than non-compliant reports. Hence, the early adopters' differentiation strategy might lead to a higher proportion of companies reporting under the GRI framework, and, thus, these companies incorporate a higher level of numerical content into their sustainability reports.

Moreover, it could also be assumed that the collection of numerical data is relatively expensive. As the net benefit of sustainability reporting for late reporters is minimal or even negative, they are unwilling to invest in numerical data collection. Therefore, these companies might seek to produce their reports in a way that minimizes costs. Hence, the differentiation strategy, as well as the cost intensity of numerical data, could be the underlying reasons for the observation.

Furthermore, the difference in horizon content between the two types is marginal but statistically significant at 1% (Table 4). The effect could potentially be explained by the survey conducted by Bhimani et al. (2016). The results suggest that early rather than late adopters tend to incorporate sustainability aspects into their three-or five-year plan as well as in their business vision. As described above, these companies have real incentives to report on their sustainability issues. For this reason, they might also include these topics in their future strategy and incorporate it into the statements mentioned above. This behavior could explain why reports from early adopters tend to contain more horizon content. In addition, there is also a small difference with regards to target orientation that reports from early adopters tend to include more target-related words. Nevertheless, this difference is not statistically significant.

Referring to the readability, the corresponding index indicates that early adopters' reports are less readable (Table 4). Nonetheless, this is statistically insignificant. Looking at the individual components, there are only minor differences in the decimal places, which are also not statistically significant. Consequently, one could assume that the different reporting types and the associated incentives have no impact on readability. However, even if there are no differences, readability is still at a very high level, which increases the difficulty for stakeholders to comprehend the content of the reports accurately.

With reference to the economic, environment, and social dimension, the reports from early adopters tend to contain more content from all three dimensions. The difference in social content is statistically significant at 1%. Even if the re-

ports from late adopters are characterized by a lower degree of social content, it is still higher than the economic and environmental content of reports from early adopters. Bhimani et al. (2016) explain that early adopters report on their sustainability responsibilities to create value, while late adopters report on these issues to deflect threats to their brand image. Hence, one explanation could be that late adopters report only the minimum amount of content to satisfy stakeholders and maintain their brand image. In contrast to the other group, they are not as intrinsically motivated and, thus, they tend to report relatively fewer contents than early adopters.

However, the difference in economic and environmental content between reports from early and late adopters is marginal and not even significant at 10%. This finding seems to contradict the previous insights. Nonetheless, the study conducted by Bhimani et al. (2016) also showed that late adopters try to keep up with their competitors by imitating them. For this purpose, they use similar reporting methods, resulting in reports with similar topics (Bhimani et al., 2016). As a result, this imitation strategy could explain why late adopters obtain a similar level of economic and environmental content, although they publish shorter reports and receive a marginal net benefit from reporting.

Referring to the number of reported GRI buckets, early reporters include an average of about sixteen buckets, while late adopters report an average of about thirteen buckets. This difference is significant at 1%. One possible explanation is based on the differentiation strategy mentioned above. Early reporters authentically pursue CSR reporting to differentiate themselves from their competitors. Therefore, they might use the opportunity to report under the GRI framework to stand out further. It turns out that relatively more reports from early adopters are in line with the framework compared to late adopters. If companies report according to the guidelines, there might be a high probability that they use the related topics as a guideline, resulting in a greater coverage of GRI topics.

To sum up, there are significant differences between the reports from early and late adopters. The current literature indicates that the reasons for those could be related to the early adopters' differentiation strategy and the underlying differences in reporting incentives.

7. Implications of CSR Performance

7.1. Empirical Model

As the data has been collected and the overall dataset, as well as the characteristics of the different groups, have been described, the formulated research hypotheses can be tested. The following empirical model is utilized to test the hypotheses:

$$
\begin{aligned}
&CSR\ Performance \\
&\quad = \beta_0 + \beta_1 LGTH + \beta_2 NC + \beta_3 HC + \beta_4 TO \\
&\quad + \beta_5 TN + \beta_6 RDB + \beta_7 ECON + \beta_8 ENV \\
&\quad + \beta_9 SOC + \beta_{10} ITMS + \beta_{11} SZ + \beta_{12} ROA \\
&\quad + \beta_{13} CURRAT + \beta_{14} CAPSP + \beta_{15} PNEW \\
&\quad + \beta_{16} LEV + \beta_{17} INST + \beta_{18} ADT + \beta_{19} GRI \\
&\quad + \beta_{20} INTG + \sum_{j=1}^{9} \mu_j Year + \sum_{k=1}^{11} \gamma_k Industry
\end{aligned}
$$

Based on this empirical model, a linear regression is conducted using R and the corresponding *caret* package. The model allows forecasting the relation between different input variables and an independent output variable.

7.1.1. Dependent Variable

The dependent variable of the regression is the individual CSR performance. Consistent with the study of Clarkson et al. (2020), the corresponding performance data was retrieved from the ASSET4 database. The performance score is an ESG score ranging between 0 (*weak performance*) and 100 (*strong performance*) and is based on 178 indicators. Thereby, it comprises the following dimensions: resource use, emission, innovation, workforce, human rights, community, product responsibility, management, shareholders, and CSR strategy (Eikon, 2017). As described in the beginning, ASSET4 offers extensive information on the sustainability aspects of companies worldwide. Moreover, researchers argue that using this data is beneficial since the underlying evaluation of companies' CSR performance is more comprehensive and objective compared to other sustainability databases (Clarkson et al., 2020). In addition, the corresponding scores are not solely based on the companies' sustainability reports, but also incorporate information from "stock exchange filings, [...] annual reports, non-governmental organizations' websites, and various news sources" (Ioannou & Serafeim, 2012, p. 21). Therefore, this underlines that the validity of the data is ensured.

In contrast to Clarkson et al. (2020), the combined ESG score of the ASSET4 database is utilized, and no dimension is excluded. This score is based on social, environmental, and corporate governance factors. Thereby, the ESG performance score reflects the overall concept of sustainability.

7.1.2. Independent Variables

The variables of interest are the textual characteristics as well as additional company- and report-specific characteristics. *LGTH* reflects the length of reports. Due to the risk of multicollinearity, only one of the length indicators is considered in the regression. For this reason, the logarithm of the number of characters serves as a proxy for the report length. *NC*, *HC*, and *TO* are the variables representing the degree of numeric content, horizon content, and target-related words

Table 4: Descriptive Statistics – Early & Late Adopter

Variable	Early Adopter	Late Adopter	Sig. Level
#Characters	323,956	136,032	1%
#Total Words	27,449	11,349	1%
#Unique Words	2,265	1,510	1%
Numeric Content	87.3	77.8	5%
Horizon Content	1.0	0.8	1%
Target Orientation	6.4	5.8	n.s.
Tone	0.006	0.005	n.s.
Readability Index	0.50	0.45	n.s.
Flesch Reading Ease	23	23	n.s.
Fog	17	17	n.s.
Flesch-Kincaid	21	21	n.s.
Economic	0.065	0.063	n.s.
Environment	0.146	0.138	n.s.
Social	0.174	0.157	1%
GRI Index	16.2	12.6	1%

Mann-Whitney-U-Test (1%, 5%, 10%, or not significant (n.s.))

per 1,000 words. On top of that, *TN* represents the tone indicator, while *RDB* is the readability index. The individual readability scores are excluded from the regression due to the risk of multicollinearity. The thematic CSR disclosure dimensions are *ECON* (economic), *ENV* (environment), and *SOC* (social). *ITMS* reflects the number of GRI topics covered per report. In addition to the textual characteristics, dummy variables are included. *ADT* is a dummy variable, whether the company is an early or late reporter. Furthermore, *GRI* reflects whether the report is in accordance with the GRI framework or not. Lastly, the dummy variable *INTG* indicates whether the report is an integrated or stand-alone report.

7.1.3. Control Variables

Control variables are included in the regression to control other factors that might influence the underlying performance. These variables comprise both financial as well as non-financial variables. The choice of control variables is based on previous studies (e.g., Clarkson et al., 2008, 2011; Nazari et al., 2017). The first control variable is *SZ*, which serves the purpose of reflecting the size of the company. For this purpose, it is measured by the natural logarithm of total assets. Another control variable is *ROA*, which is the income of the enterprise divided by its total assets. *CURRAT* is measured by the ratio of total current assets to total current liabilities. *CAPSP*, which reflects the degree of capital spending, is the ratio of total spending to total sales revenues. Moreover, *PNEW* is the amount of net property, plant, and equipment divided by gross property. *LEV*, which reflects the company's leverage ratio, is calculated by dividing total debt by total assets.

Furthermore, *INST* is the percentage of institutional ownership. Since the database does not provide such information, other measures were used as a proxy for institutional ownership. As the database provides information on the percentage of strategic ownership differentiated by different owner types, INST equals the sum of strategic ownership held by institutions, investment banks, and pension/endowment funds. All of these data points were retrieved from the WorldScope databank via the financial information service of Thomson Reuters. WorldScope provides fundamental financial data on leading companies worldwide. For each year, the value at the end of the year was taken to ensure consistency.

Lastly, dummy variables for the industry and year were added. These variables control for specific temporal and industry effects. Nonetheless, the database does not include the corresponding control data for all companies in all periods. Thus, the observations with missing control variables were excluded from the regression. This procedure led to the exclusion of 532 reports resulting in a final regression sample of 1,899 reports.

7.2. Assumptions of Linear Regression

Before one can interpret the results, the four assumptions of a linear regression have to be tested. These assumptions include linearity (linear relationship), normality, homoscedasticity, and the avoidance of multicollinearity. The corresponding tests will be conducted with the statistical software R and are mainly based on the visualization of the data. If one of these assumptions does not hold, the scientific findings of the regression are inefficient or even strongly biased. Hence, the regression assumptions are necessary to accurately interpret the results (Statistics Solutions, n.d).

Firstly, multicollinearity means that at least two independent variables are highly correlated and demonstrate a linear relationship. Hence, this causes that one of the different

independent variables is redundant. The occurrence of multicollinearity has to be avoided since it leads to somewhat unstable parameter estimates. There are various methods to test for multicollinearity. In this thesis, the Variance Inflation Factor (VIF) is utilized. Multicollinearity causes that the variance of the regression increases and, hence, makes it unreliable. The VIF makes use of this phenomenon and measures how much of the inflated variance is due to multicollinearity. A corresponding value above four indicates that multicollinearity might exist and that a further analysis should be conducted. If the VIF score is higher than ten, this is a strong indication for multicollinearity. In this case, this has to be corrected (Corporate Finance Institute, n.d). In Appendix 35, the different VIF values for each predictor are shown. The indicator for length, which is the logarithmic number of characters, and the number of reported GRI buckets obtain the highest VIF values. These values are slightly above four. If one of the variables is excluded from the regression, the VIF values tend to be lower. This insight indicates that these predictors are somewhat correlated. This correlation is evident because when companies report on more GRI topics, the reports inevitably become longer. Since the VIF values are above the limit of four, this indicates that multicollinearity might exist. Thus, the variable representing the number of GRI items reported is excluded from the regression to solve this problem. In turn, the VIF value of report length decreases to around 2.5. The new values can be seen in Appendix 36.

Secondly, the second assumption, which is linearity, means that the relationship between the dependent variable and the numerous predictors is linear. This assumption implies that the outcome variable is a straight-line function of the different predictors. Moreover, it indicates that the regression slope does not depend on the value of the other independent variables. To test this assumption, one can plot the residuals against the fitted values. The assumption holds if one can detect a horizontal line with no specific pattern (Nau, n.d). The corresponding plot can be seen in Appendix 37, which provides evidence that a linear relationship between the numerous predictors and the CSR performance exists.

Thirdly, the assumption of normality means that the residual errors are assumed to be normally distributed. A violation of this assumption has a negative effect on the significance level of the coefficients and the corresponding confidence intervals. This assumption can be tested using a QQ-plot with the theoretical quantiles on the x-axis and the standardized residuals on the y-axis (Nau, n.d). This plot is shown in Appendix 38. Since the data is on the diagonal dashed line, the data is assumed to be normally distributed. Hence, the normality assumption also holds.

Fourthly, the last assumption is the presence of homoscedasticity. Homoscedasticity means that the error terms of all values of the predictors are on the same level. If the assumption does not hold and heteroscedasticity is present, this leads to biased standard errors. In turn, this results in false conclusions about the significance of the numerous regression coefficients. To test for homoscedasticity, one can use the scale-location plot, which shows the fitted values on the x-axis and the root of standardized errors on the y-axis (Statistics Solutions, 2013). This plot is shown in Appendix 39. It demonstrates a horizontal line with a small downward tendency on the right side. This shows that the associated points tend to be equally spread. Hence, the fitted values' variances tend to be constant among all values except a small deviation on the right side. Therefore, it can be assumed that the assumption of homoscedasticity holds.

Moreover, besides the analysis of the regression assumptions, one should look at the existence of outliers. With regard to the dataset of this study, outliers are reports with strongly deviating characteristics. Using the statistical software R, these values are highlighted in the different plots. Three outliers can be identified. In the next step, one should check how realistic the values of these outliers are. For this purpose, the values are compared to the median values of the overall dataset. The textual characteristics of the three outliers identified seem to be reliable. However, their ESG scores are relatively low since these values are close to or even zero. For this reason, one could consider excluding the outliers from the dataset. However, as the ASSET4 database publishes these values, the three corresponding reports will be retained in the analysis. Nonetheless, a more in-depth analysis concerning outliers will be conducted in the robustness section.

In summary, the underlying assumptions of a linear regression are fulfilled. In addition, a few outliers are identified. Since the outliers and the corresponding characteristics seem to fit the dataset except for the CSR performance deviations, the reports remain in the dataset. However, this will be further considered during the robustness tests of the regression. All in all, the regression results can be interpreted since the underlying regression assumptions are fulfilled.

8. Results

8.1. Hypotheses Testing

After conducting the textual analysis, collecting further data, and testing the regression assumptions, the regression can be performed, and its results can be interpreted. Before the individual hypotheses are analyzed, the explanatory power of the regression is observed. The adjusted R^2, which adjusts for the number of independent variables, amounts to 0.551. Hence, 55.1% of the variability of the CSR performance is explained by the different predictors. The study of Patten (2002) examined the relationship between environmental disclosure and the corresponding performance, and its model obtained an adjusted R^2 of 0.38. The researcher states that the "explanatory power is relatively high" (Patten, 2002, p. 770). In contrast, the study of Clarkson et al. (2020) observed a similar relation between sustainability performance and disclosure and obtained an adjusted R^2 of around 0.72. Therefore, the corresponding value of this study is between the values of the current literature. This comparison emphasizes that the explanatory power of this

study is acceptable. Moreover, the F-statistic indicates that the regression model is statistically significant at 0.1%. The regression table is shown in Table 5.

The first hypothesis predicts that the length of sustainability reports is positively correlated with CSR performance. The regression coefficient is positive and statistically significant at 0.1%. This coefficient implies that companies obtaining higher sustainability performance tend to produce more comprehensive sustainability reports. This insight is consistent with the finding of Clarkson et al. (2020) that companies with superior CSR performance prepare longer reports. The underlying reason could be the fact that they can report on more topics due to a larger number of pursued CSR activities. Moreover, this is also in line with the voluntary disclosure theory and contradicts the legitimacy theory. All in all, the hypothesis about the relationship between report length and sustainability performance is supported.

The second hypothesis states that the readability of CSR reports negatively correlates with the corresponding performance. This hypothesis means that reports from well-performing companies are assumed to be more readable. Looking at the associated regression coefficient, it contradicts the hypothesis. The coefficient implies that companies that prepare reports with a low level of readability have a higher CSR performance. This finding is consistent with the study of Clarkson et al. (2020), although the researchers expected a different relationship. The researchers argue that more sophisticated analyses might cause poor readability. However, the coefficient is not statistically significant. To sum up, the regression analysis does not support the second hypothesis.

The third hypothesis forecasts that the tone of sustainability reports is negatively correlated with the associated performance. However, the regression generated a positive coefficient, which is statistically significant at the 1% level. This outcome indicates that companies with a poor (good) CSR performance publish reports with a more negative (positive) tone. As described in the literature review, this can be explained by the behavior of managers of poor performing companies, who include more negative phrases to fend off criticism and/or build up credibility (Clarkson et al., 2020). For these reasons, they might use a more negative tone in their sustainability reports. In summary, the third hypothesis is not supported.

Hypothesis 4a predicts a positive relationship between the number of reported GRI items and CSR performance. Due to multicollinearity, this variable was excluded from the regression. However, in order to examine this hypothesis, the length indicator was excluded from the regression, and the indicator for the number of GRI items was included. In this way, the occurrence of multicollinearity is avoided. The remaining regression assumptions are also tested for this regression equation. The regression reveals a positive coefficient of around 0.49, which is significant at 0.1%. This finding supports the hypothesis that companies, which tend to report on more GRI topics, have a higher CSR performance.

Hypothesis 4b states that topic-specific CSR disclosure positively correlates with sustainability performance. Referring to the environmental and social dimensions, the regression results imply that the relationship between these dimensions and the performance is positive. The coefficient of environmental disclosure is significant at 0.1%, while the social disclosure coefficient is significant at 1%. This is in line with the voluntary disclosure theory and the findings of the current literature (e.g., Al-Tuwaijri et al., 2004; Clarkson et al., 2008). In contrast, the coefficient of the economic dimension is negative, which contradicts the hypothesis. One could argue that this is based on the fact that the dependent variable is an ESG score, which does not consider economic factors (Eikon, 2017). Moreover, the associated coefficient is not statistically significant. As the economic dimension is not part of the ESG score, this might explain that the degree of economic disclosure has no significant effect on the dependent variable. In conclusion, the hypothesis is widely supported, except for the relationship between economic disclosure and CSR performance.

The fifth hypothesis predicts that companies with superior CSR performance tend to include more numerical data into their reports. The related coefficient is negative, but close to zero (-0.002). As shown in Table 1, the difference between the minimum and maximum numeric content for the overall dataset is around 450 numbers per 1,000 words. According to the regression, this difference in the extreme values would result in a 0.72 difference in CSR performance. This underlines that the impact is marginal. Moreover, the coefficient is not statistically significant. Therefore, the stated hypothesis is not supported by the regression analysis.

Moreover, the sixth hypothesis forecasts a positive relationship between the degree of horizon content and sustainability performance. However, the corresponding coefficient indicates that the relationship is the other way around. This deviation might be caused by poor performing companies that focus on future performance to distract from the current one. These companies might also incorporate future-related content to emphasize their ambition to improve their performance in the future (Asay et al., 2018). Nonetheless, the coefficient is not statistically significant. Thus, the hypothesis is not supported.

The seventh hypothesis deals with the relationship between the degree of target orientation and CSR performance. Based on the goal-setting theory and the voluntary disclosure theory, it was assumed that firms with a superior CSR performance tend to formulate more targets in their sustainability reports (Latham & Locke, 1979). The former allows managers to motivate employees and direct their behavior. In contrast, the latter allows these firms to show the market that they are able to meet their stated objectives. As a result, they incorporate more target-related words into their sustainability reports. However, a necessary condition is that the formulated objectives are clear, specific, and attainable (Latham & Locke, 1979). The regression results confirm this hypothesis since the corresponding coefficient is positive. Furthermore, it is statistically significant at 0.1%. In summary, the

hypothesis about the relation between the degree of target orientation and CSR performance is confirmed.

As the hypotheses concerning the textual characteristics have been analyzed, the hypotheses concerning the other characteristics will be observed in the next step. The eighth hypothesis is related to whether a company is defined as an early adopter or a late adopter. The associated coefficient shows that being a late adopter has a negative effect on sustainability performance. According to the regression, this effect amounts to around 9.8 performance points. Considering that the ESG score ranges between 0 and 100, this is an enormous difference. In addition, the coefficient is significant at the 0.1% level. This outcome is consistent with the study of Bhimani et al. (2016), which states that late adopters are only motivated to pursue sustainability reporting to improve their reputation. Thus, late adopters only want to keep up with their competitors and do not authentically pursue CSR activities. In turn, these companies tend to obtain a lower CSR performance. The results are also in line with the voluntary disclosure theory. To sum up, the hypothesis about the relationship between the adopter type and CSR performance is confirmed.

Concerning the hypothesis dealing with the impact of reporting under the GRI framework, the results reveal that non-complying companies, on average, obtain a score that is around 5.8 performance points lower. The coefficient is significant from zero at the 0.1% level. This outcome contradicts the study of Bernard et al. (2015) and the legitimacy theory. On the contrary, the results emphasize that companies with a superior sustainability performance commit themselves to the GRI framework to differentiate themselves from the competition. However, the stated hypothesis does not determine a specific direction but merely assumes that there is a certain relationship. Therefore, the hypothesis is supported. In addition, the impact of GRI reporting seems to be positive.

The last hypothesis states that the firm's sustainability performance has an impact on the reporting method. Nonetheless, it does not predict a specific direction. The regression analysis shows that the coefficient of preparing a stand-alone sustainability report is positive. The coefficient is statistically significant at 1%. This finding demonstrates that companies that have superior sustainability performance tend to publish stand-alone reports. The underlying reason might be that these firms want to highlight their superior performance. Therefore, they tend to publish a report that solely focuses on this aspect. In contrast, firms with poor performance might publish integrated reports to hide their sustainability failure in their financial results. To sum up, the hypothesis is supported. Moreover, the effect of integrated reporting seems to be negative.

All in all, the regression analysis confirms the hypotheses concerning length, number of GRI items, target orientation, adopter type, GRI framework, and the reporting method. Moreover, the hypothesis about the degree of topic-specific CSR disclosure is widely supported. The coefficients relating to numerical content, horizon content, and readability are insignificant, while contrary to expectations, the tone positively correlates with CSR performance. It is noteworthy that most of the hypotheses can be explained by the voluntary disclosure theory and not by the legitimacy theory. These insights underline that firms with superior CSR performance tend to be more engaged in sustainability reporting. As a result, these companies prepare, on average, longer stand-alone reports with a high degree of formulated targets and a high degree of topic-specific disclosure. In addition, these reports tend to be in accordance with the GRI framework and cover many of the related topics. In contrast, underperforming companies tend to prepare short reports integrated into their annual reports and contain a few formulated targets. Moreover, these reports are rarely prepared using the GRI framework.

8.2. Robustness Tests

Different robustness tests are performed to check the strength of the regression model. These tests should confirm the model as well as the corresponding findings by applying different conditions. The first robustness test refers to the definition of early and late adopters. As described in the literature review, late adopters are defined as companies that have started reporting in response to the announcement of the EU Directive in 2014. However, according to Fiechter et al. (2019), some EU companies might have anticipated the regulation and started reporting before the announcement. Therefore, in the first robustness test, late adopters are defined as companies that started reporting in 2013 or later. Thereby, the number of reports from late adopters increases from 58 to 169 reports. The significance levels of the individual predictors before and after the robustness test, as well as the corresponding coefficients, can be seen in Appendix 40.

For most predictors, the results remain unchanged. While the significance level of the tone value coefficient increased from 1% to 0.1%, the significance level of the coefficient for publishing an integrated report decreased from 1% to 5%. However, the results remain constant and are only marginally affected. In addition, the coefficient of horizon content becomes slightly significant at 10%. As the correlation is negative, companies with a poor CSR performance focus more on future aspects in their reports. The underlying reasons might be that these companies want to distract from the past or positively shape stakeholders' perception by emphasizing future initiatives (Asay et al., 2018). This outcome contradicts the originally formulated hypothesis about the relationship between CSR performance and horizon content. However, this hypothesis was not confirmed in the main analysis, either. All in all, a different definition of early and late adopters leads to only marginal differences in the various variables and, hence, confirms the previous findings.

The second robustness test involves another dependent variable instead of the ASSET4 ESG score. Other studies use different sustainability scores, such as the performance data from Kinder, Lydenberg, and Domini (KLD) or the Bloomberg ESG Score (e.g., Clarkson et al., 2020; Nazari et al., 2017).

Table 5: Regression Results

Variable	Coefficient	Relationship	Sig. Level
LGTH	10.100	+	0.1%
NC	-0.002	-	n.s.
HC	-0.776	-	n.s.
TO	0.481	+	0.1%
TN	133.274	+	1%
RDB	0.431	+	n.s.
ECON	-9.339	-	n.s.
ENV	42.929	+	0.1%
SOC	18.742	+	1%
ITMS[1]	0.488	+	0.1%
GRI (No)	-5.823	-	0.1%
ADT (Late)	-9.832	-	0.1%
INTG (No)	1.747	+	1%
SZ	3.205	+	0.1%
ROA	0.079	+	5%
CURRAT	-1.113	-	0.1%
CAPSP	-0.018	-	n.s.
PNEW	-8.456	-	0.1%
LEV	9.827	+	0.1%
INST	7.270	+	10%
Temporal Effects		Yes	
Industry Effects		Yes	

1 = Separate regression without LGTH due to multicollinearity

However, this data is not available in the database of Thomson Reuters. Thus, the ESG score from the ASSET4 database is adjusted. Following the approach of existing studies (e.g., Clarkson et al., 2020; Ioannou & Serafeim, 2012), the governance dimension is excluded from the overall score. In the ASSET4 database, the individual scores for each ESG dimension are available. The average score of the environmental and social dimensions is computed to exclude the governance dimension. This score acts as a new dependent variable to test the robustness.

Looking at the results in Appendix 41, the coefficient of the tone variable became insignificant. However, in the main analysis, the corresponding hypothesis was not supported either. Furthermore, the significance of social disclosure's coefficient decreased from 1% to 5%, while the negative coefficient of economic disclosure became significant at 1%. However, the hypothesis regarding the latter coefficient was not supported before. More interesting is the change in the readability coefficient. In the main analysis, the coefficient was not significant and positive. In this robustness test, the coefficient became negative and significant. This outcome supports the stated hypothesis that firms with poor CSR performance produce less readable reports. An underlying reason might be that they aim to obfuscate their poor CSR performance. Lastly, the effect of producing an integrated report became insignificant. This change contradicts one of the critical findings of the main analysis. To sum up, the main findings remain constant except for the impact of producing an integrated report.

The third robustness test manages the occurrence of outliers. To identify all outliers, the approach of John Tukey, a researcher who invented the boxplot, is applied. For this purpose, the interquartile range (IQR) has to be calculated. The IQR is the difference between the 25^{th} and 75^{th} percentile. All data points that are one and a half times the IQR higher (lower) than the 75^{th} percentile (25^{th} percentile) are defined as outliers (Purplemath, n.d). The outlier identification process is performed for each numeric variable. Instead of removing these values from the dataset, values outside the lower limit are replaced by the value of the 5^{th} percentile. Moreover, all values that lie outside the upper limit are replaced by the corresponding value of the 95^{th} percentile. In this way, the dataset is cleared for outliers. The new levels of significance, as well as the new coefficients, are listed in Appendix 42.

In addition to various changes in the control variables, the coefficient of numeric content becomes significant at 1%. The results indicate that there is a negative relationship between this variable and CSR performance. This finding demonstrates that companies with poor CSR performance include more numerical data in their reports. Even if a positive relationship was predicted, the hypothesis was not supported in the main analysis. Furthermore, the significance levels from tone and target orientation decrease slightly. All

in all, the results remained mostly unchanged.

The fourth robustness test is related to future and past sustainability performance. The analysis examined the relationship between textual characteristics and CSR performance in the same period. However, one could argue that companies have an excellent performance in period $t-1$ but incorporate this into their reports for period t. A potential explanation could be that they were not aware of their excellent performance. This phenomenon might be particularly prevalent if, for example, they publish the report shortly after the end of the financial year. In this case, they might not be aware of their excellent CSR performance and incorporate this into the subsequent report. This might lead to a time-delayed effect. Moreover, one could also imagine a similar effect in the opposite direction. For example, due to upcoming initiatives, firms are aware that they will obtain a superior CSR performance in the period $t+1$. For this reason, they design their reports differently in period *t,* highlighting their upcoming performance. Therefore, robustness tests that examine the relationship between textual characteristics and future (past) performance are conducted.

The corresponding regression tables are shown in Appendix 43 and Appendix 44. Most of the changes are only marginal. The only remarkable effect concerns the social disclosure coefficient. This coefficient becomes insignificant if past performance acts as the dependent variable. However, the remaining coefficients remain mostly constant.

For the fifth robustness test, an additional control variable that captures the reporting experience is introduced. The variable *EXP* reflects the number of reports under the GRI framework since 2005. For instance, if a company has published eight reports from which six reports are according to the GRI framework since 2005, the corresponding value is equal to six. All reports before 2005 are not relevant for this analysis. The regression results are listed in Appendix 45. There are only minor changes in the significance levels, but the key insights remain robust. The coefficient *EXP* is positive and significant at 1%, indicating that reporting experience positively affects CSR performance.

Finally, following the approach of Mittelbach-Hoermanseder et al. (2019), an additional control variable was introduced taking into account the effect of national culture. This aspect is particularly important since some countries, such as France, Denmark, and the UK, have already introduced regulations on the disclosure of sustainability information before the announcement of the EU Directive (Fiechter et al., 2019; Hummel & Rötzel, 2019). This could have an impact on the results. The first two letters of the International Securities Identification Number (ISIN) indicate the company's country. These letters were extracted and used as an additional dummy variable to account for national differences.

After performing the regression with the new control variable, the results demonstrate some interesting changes. These can be seen in Appendix 46. Similar to the first robustness test, the coefficient of the horizon content became significant. This result contradicts the initial hypothesis and supports the theory that CSR performance and horizon content are negatively correlated. Moreover, the coefficient of publishing an integrated report became insignificant, similar to the second robustness test. The other variables remained mostly unchanged. Therefore, this robustness test confirms the key findings of this thesis illustrated previously except for the relationship between report type and sustainability performance.

In summary, the four robustness tests caused only marginal changes for the different variables. Most of the changes concerned variables whose hypothesis had not been confirmed before. Nevertheless, the hypothesis regarding the report type was not always supported. When choosing a different ESG score or controlling for national differences, the coefficient became insignificant. These robustness tests underline that this variable is sensitive to the conditions and question the relationship between this variable and CSR performance. Nonetheless, the robustness tests underline that the overall model is largely robust, and the main findings remain mostly unchanged.

8.3. Additional Analysis

So far, the study confirms the hypotheses concerning the relationship between CSR performance and report length, environmental content, social content, number of GRI items covered, target orientation, adopter type, GRI framework, and reporting method. Robustness tests in the previous sections also supported these hypotheses. However, the causal direction of the regression goes from textual characteristics to CSR performance. This structure allows to analyze the relationship between textual characteristics and CSR performance in a single regression. Nevertheless, the underlying reasoning of the hypotheses is based on the effect of CSR performance on textual characteristics. However, one could argue that regression results only reflect a correlation and not a causal relationship. This argumentation would mean that the underlying regression is valid.

However, additional analyses are performed. These are intended to show that the findings are valid with CSR performance as the independent variable and text characteristics as the dependent variables. Therefore, a regression is performed for each variable of the confirmed hypotheses. This approach results in a total of eight additional regression equations. The regression structure is similar to the previous analysis. The control variables remain the same. The eight text characteristics are each used as dependent variable, while CSR performance is the predictor. In addition, the dummy variables concerning the report type, the GRI framework, and the adopter type are also included as control variables. The individual variables are not included if they are used as the dependent variable themselves. In this case, the variable is omitted because, otherwise, it would act as both a dependent and an independent variable.

Concerning report length, target orientation, number of GRI items, and environmental and social disclosure, a linear regression is performed. However, this is not possible for the other three variables, as they are binary variables. In this

case, a logistic regression is performed that is capable of modeling a binary dependent variable. The regression equation, for example, for the report length, is as follows:

$$\begin{aligned} LGTH = \beta_0 + \beta_1 CSR\,Performance + \beta_2 SZ + \beta_3 ROA \\ + \beta_4 CURRAT + \beta_5 CAPSP + \beta_6 PNEW + \beta_7 LEV \\ + \beta_8 INST + \beta_9 ADT + \beta_{10} GRI + \beta_{11} INTG \\ + \sum_{j=1}^{9} \mu_j Year + \sum_{k=1}^{11} \gamma_k Industry \end{aligned}$$

The regression tables are shown in Appendix 47 – Appendix 54. For each regression, the coefficient for CSR performance is significant at 1% or even 0.1%. Furthermore, the direction of the coefficient in each regression is consistent with the hypotheses. For instance, the CSR performance coefficient is negative when the GRI variable is the dependent variable. A regression outcome close to one predicts that the report is not in accordance with the GRI framework. Hence, this outcome shows that the higher the ESG score is, the more the model predicts that the corresponding report is in line with the GRI framework. This prediction is consistent with the hypothesis that firms with superior CSR performance tend to report under this framework. All in all, the additional analyses confirm the previous findings.

8.4. Managerial Implications

The findings of the current analysis provide manifold implications for businesses, regulators, analysts, and other market participants. Firstly, the study shows that narrative characteristics (e.g., readability, report length) and additional characteristics (e.g., adopter type, GRI compliance) allow stakeholders to assess businesses' sustainability performance. The associated performance is an essential factor for the decision-making process of consumers and investors. As a result, managers have to consider this during the creation process of their sustainability report. Instead of only considering which contents they include in their reports, managers also have to consider how they communicate the corresponding contents. This approach allows managers to actively shape stakeholders' perception and convince them of their sustainability efforts. Reporting under the GRI framework or incorporating a high degree of environmental content, for example, can lead to CSR performance being rated higher than it genuinely is. If they do not consider these aspects, stakeholders might perceive the firm's sustainability performance differently.

Hence, managers should invest more time and resources into the preparation process of their sustainability reports to optimally design the textual narratives. Managers could set up a dedicated department as well as internal guidelines to ensure a sufficient reporting quality. As a result, the creation of such reports receives full attention. This investment guarantees that the textual design is optimized. To check the format and the quality of their sustainability reports, firms could utilize textual analysis. Thereby, they could analyze their past reports, which helps them to identify their weaknesses. Afterwards, they can tackle these weaknesses and publish an improved report for the next period.

Furthermore, companies should design their reports according to the findings of the study. One possibility would be to commit themselves to the GRI framework voluntarily. As shown in the regression analysis, this signals market participants a high level of sustainability performance. Even if they cannot deliver this level of performance, they can positively shape their stakeholders' perceptions.

Moreover, in particular for private investors, it is often challenging to assess the actual sustainability performance of companies. This is based on the fact that the access to databases such as the ASSET4 is often subject to a fee. Hence, investors often have to rely on the contents of sustainability reports to assess companies' CSR performance. However, the study expands the toolkit of private investors since they can utilize the linguistic features as an additional proxy for a firm's sustainability performance. Since the latter factor is an increasingly important decision-making criterion, this offers an immense benefit for private investors. Furthermore, this is also a vast advantage for institutional investors and analysts. Different databases do not always cover small and private companies. Thus, these investors also have difficulties assessing the actual CSR performance and can use linguistic features as an additional dimension for their analysis.

Lastly, as shown in chapter six, *Descriptive Statistics*, there are substantial differences in the design of sustainability reports. While some companies publish only a few pages within their annual report, other companies intensively describe their sustainability efforts substantiated by numerical data and in accordance with the GRI framework. Even if stakeholders can derive the corresponding performance from such characteristics, it is challenging to compare the reports among different companies. This is a major finding for regulators in Europe. Even if the EU Directive mandates that companies report on their sustainability issues, companies are not obliged to use a specific framework. This increases the pressure on regulators to tighten the regulatory framework to make sustainability reporting more comparable and consistent. Thereby, they should consider following the approach of financial reporting and introducing a specific reporting framework.

8.5. Theoretical Implications

From a theoretical point of view, this thesis contributes to the literature in several respects. Firstly, it is the first study examining the differences between integrated and stand-alone reports as well as the differences between reports from early and late adopters by conducting a textual analysis. While many researchers focused solely on integrated or stand-alone reports (e.g., Clarkson et al., 2020; Muslu et al., 2019; Nazari et al., 2017), no prior study already investigated the differences between these reports and explained the underlying reasons. Referring to integrated and stand-alone reports, this study explores significant differences, for instance, concerning report length, content, and readability. Moreover, the re-

sults also reveal significant differences between the reports from early and late adopters.

Secondly, this study examines textual characteristics, which have not been observed by other researchers. Mittelbach-Hoermanseder et al. (2019) measured topic-specific CSR disclosure by the cosine similarity. However, they measured the topics of the EU Directive, while this study measures the three dimensions of the GRI Standards. In addition, another textual feature that was previously unobserved is the degree of target orientation. Targets are a fundamental element in the field of management accounting and control. Therefore, it is an interesting finding that target orientation also plays an essential role in sustainability reporting. This study is also the first to utilize textual analysis to approximate a GRI index for each report. This approach allows to determine how many topics are covered, even if the reports are not following the GRI framework. To sum up, the results imply that target orientation, topic-specific CSR disclosure, as well as the number of GRI topics covered, significantly correlate with CSR performance.

Thirdly, the study is one of the first observing the impact of the EU Directive entering into force in 2017. The observation period allows to detect any changes in response to the shift to mandatory sustainability reporting. Since the shift happened in 2017, previous literature could not examine this impact. Therefore, they could only observe the effect of the announcement of the EU Directive in 2014, which was not yet binding this year. The results imply that more GRI topics have been covered after the regulation came into force (Appendix 23).

Fourthly, some studies also examined the relationship between textual features of sustainability reports and the associated performance (e.g., Clarkson et al., 2020; Hummel & Schlick, 2016; Nazari et al., 2017; Patten, 2002). However, this study focuses not only on textual features but also on other characteristics such as the report type (integrated or stand-alone), adopter type, and the use of the GRI reporting framework. The results of the regression imply that all of these characteristics significantly correlate with sustainability performance, which display relevant insights expanding current literature.

Moreover, this study also has some implications for future research. Since the literature about the companies' motivations for preparing an integrated or stand-alone report is limited, future research can build on this research opportunity. These insights would enable researchers to precisely explain the differences between integrated and stand-alone sustainability reports. Furthermore, future research can investigate the differences between reports from voluntary adopters and resisters. The latter group postponed the reporting process until the EU Directive came into force in 2017. Due to the observation period of the study until 2018, this grouping could not be observed separately. Lastly, since the sample is based on the STOXX Europe 600, the relation between sustainability disclosure and the corresponding performance in other regions might be an interesting subject for future research. Since the analysis is based on a code in R, it can be performed easily and time efficiently for other regions. To do so, only the reports to be analyzed have to be collected. Therefore, the same study could be replicated, for example, with reports from US companies to discover regional differences.

8.6. Limitations

This thesis also has some limitations, which need to be considered. Firstly, the number of reports from late adopters is relatively small. Therefore, one could question the representativeness of this group. While the sample comprises 2,373 reports from early adopters, there are only 58 reports from late adopters. The underlying reason for this is the utilized sampling rule, which states that only reports from companies with a minimum number of four reports are considered in the study. This limit ensures that companies with only one or two reports do not distort the results in the corresponding years. Since the emergence of late adopters was initiated by the EU Directive's announcement in 2014, late adopters could have published a maximum of five reports covering the fiscal years between 2014 and 2018. Late adopters, which started reporting even later in 2016, were excluded from the regression due to the sampling rule mentioned above.

In addition, as described in the methodology part, individual reports were sometimes not accessible due to various problems. In some cases, for instance, the reports were removed from the corporate website, or only a web version was available. Hence, a single missing report from a late adopter could lead to the exclusion of all other reports from that company, as they could have published a maximum of five reports during that period. Consequently, the minimum number of reports per company could have been reduced to counteract this problem. However, this would have called into question the reliability of the entire dataset. For this reason, it was decided to retain the sampling rule. As a result, the small sample size of late adopters may not be truly representative of the entire group of late adopters.

Secondly, both the Fog and Flesch-Kincaid readability values appear to be above the indices' regular score range. This deviation makes it more challenging to interpret the corresponding values accurately. Both indices refer to the required years of education. Therefore, index values above twenty seem difficult to be interpreted since this number of years of education is rather unlikely. However, the Flesch Reading Ease scores are within their normal value range. This index score is based on the same textual components as the other indices. The calculation of all three indices is based on the number of syllables per word and the number of words per sentence (Li, 2008). This shows that the data on which the calculation is based has to be correct since the Flesch Reading Ease, as well as the results of all other textual characteristics, appear plausible. For this reason, the results of the Flesch-Kincaid index, as well as the Fog index, have to be valid. Therefore, a limitation of the present study is the difficulty to interpret the corresponding results because they exceed the indices' regular range. Nonetheless, conclusion can still be drawn from the differences in readability.

Thirdly, some companies, which report under the GRI framework, publish the GRI index separately and do not incorporate it into their sustainability report. Hence, this can lead to deviations in the results of the textual analysis. The GRI indices contain additional information on sustainability topics and also comprise different required KPIs. Thus, this can impact the different textual variables, such as the indicators for thematic disclosure and numeric content. Moreover, this might have implications for the approximated GRI index. The derivation process of the GRI index is based on specific search terms. These terms are included in the text as well as in the index itself. Hence, the separate publication of the GRI index could result in some topics not being recognized. Nevertheless, this only applies to less than 1% of the sample.

Fourthly, another limitation is the lack of current literature. There is only limited literature on the impact of the EU Directive on the various textual characteristics. This problem has been solved by referring to literature examining the effects of similar regulations in the UK and France. In addition, no literature deals with the differences between integrated and stand-alone reports. Most of the papers focuses on only one of these two reporting types. This problem is further enhanced by the fact that the literature presents varying definitions of the concept of integrated reporting. While some researchers define integrated reports as annual reports with a CSR section, other researchers define them as reports in accordance with the integrated reporting framework. Due to the lack of literature, the argumentation is mainly based on theories, such as the legitimacy and voluntary disclosure theory.

Regarding the reports from early and late adopters, conclusions could be drawn from literature on the introduction of mandatory IFRS reporting. In this setting, researchers have already studied the differences between early and late adopters, as well as their motivations and characteristics. However, the extent to which there are similarities and differences between sustainability and financial reporting is questionable. In addition, when mandatory IFRS reporting was introduced, textual analysis was not as advanced as it is today. As a result, researchers have often observed features other than the textual characteristics of this study.

Fifthly, the argumentation is often based on the assumption that companies, which have a superior (poor) CSR performance, design their reports differently. Hence, the causal direction goes from the CSR performance to the different characteristics. However, one could also argue that the causal direction is the other way around. For example, this study argues that companies with a superior CSR performance report under the GRI framework to differentiate themselves. Nonetheless, one could also argue that companies, which report under the GRI framework, are able to reach a higher CSR performance since the framework allows them to structurally pursue their CSR activities. Therefore, reverse causality cannot be completely ruled out.

Lastly, another limitation concerns the sample of integrated reports. While stand-alone sustainability reports were retrieved as a whole, only the sustainability sections from integrated reports were considered for this study. For this purpose, the reports had to have a clear sustainability section, which can be extracted. Integrated reports, where the CSR information was distributed throughout the entire report, were excluded from the regression. Nonetheless, even for reports with a clear sustainability section, it is still possible that companies have also reported on sustainability issues in one of the other chapters. In turn, this might have influenced the results.

9. Conclusion

More and more companies have started reporting on their sustainability issues. The underlying reasons are manifold. A milestone in Europe was the EU's announcement in 2014 that individual companies are obliged to report on their sustainability issues from 2017 onwards. Due to the high number of reporting guidelines and the weak legal framework, companies can create their sustainability reports individually. This results in a great variety of sustainability reports. For this reason, this offered an interesting and unique research opportunity.

This study examined the relationship between sustainability disclosure and the underlying performance. Sustainability disclosure was represented by numerous variables such as the length of the reports as well as the corresponding readability. It was hypothesized that longer reports, which are easily readable with a more negative tone covering many topics and having a high degree of numeric, horizon, and target content, indicate a superior sustainability performance. Moreover, it was predicted that early adopters obtain a higher CSR performance. Simultaneously, the relationship with reporting under the GRI framework and issuing a stand-alone report was unclear.

The study was conducted for all STOXX Europe 600 companies, for which the sustainability reports were available. Besides stand-alone reports, the sample also included integrated reports that were embedded in the annual report. A textual analysis was performed using the statistical program R to preserve the textual features of the numerous reports. This methodology allows to easily replicate the study with another sample, for example, from another region.

In the first part, descriptive statistics were shown. Concerning the development over time, the results indicate that the announcement of the EU Directive led to an increase in length, target orientation, and topic-specific disclosure. In addition, readability became more difficult. Furthermore, three years later, an increase in covered GRI topics could be observed when the regulation entered into force. Moreover, integrated and stand-alone reports were identified to differ significantly from each other. Stand-alone reports tend to be longer, contain more numerical data, and include more targets, while integrated reports are less readable and tend to incorporate less content. Referring to the reports from early and late adopters, the results also revealed significant differences. The reports from early adopters tend to be longer and contain more numerical, horizon, and topic-specific content.

The underlying argumentation for these differences is based on the differentiation strategy of early adopters and the differences in reporting incentives.

In the second part, a regression was performed to examine the relationship between CSR disclosure and the associated performance. The hypotheses concerning numerical content, horizon content, tone, and readability were not supported. However, the results confirmed the hypothesis that companies with a superior sustainability performance produce longer stand-alone reports with a higher degree of target orientation and a high number of GRI topics. Moreover, the results supported the hypotheses that these companies tend to be early adopters and report under the GRI framework. The prediction concerning the topic-specific disclosure was supported to a great extent except for the relation between economic content and performance. The voluntary disclosure theory can explain most of these correlations, stating that well-performing companies voluntarily disclosure sustainability information to diminish information asymmetry between the company and investors and demonstrate its extraordinary performance.

All in all, the quantitative analysis offers various insights into the field of sustainability reporting. On the one hand, it underlines that there are significant differences between the sustainability reports of companies. On the other hand, it highlights that the expression of several textual features correlates with the underlying sustainability performance. In turn, these insights provide essential implications for academics, investors, regulators, and other market participants.

References

Al-Tuwaijri, S. A., Christensen, T. E., & Hughes, K. (2004, jul). The relations among environmental disclosure, environmental performance, and economic performance: a simultaneous equations approach. *Accounting, Organizations and Society*, *29*(5-6), 447–471. doi: 10.1016/s0361-3682(03)00032-1

Arnold, M. C., & Artz, M. (2015, jan). Target difficulty, target flexibility, and firm performance: Evidence from business units' targets. *Accounting, Organizations and Society*, *40*, 61–77. doi: 10.1016/j.aos.2014.12.002

Asay, H. S., Libby, R., & Rennekamp, K. (2018, apr). Firm performance, reporting goals, and language choices in narrative disclosures. *Journal of Accounting and Economics*, *65*(2-3), 380–398. doi: 10.1016/j.jacceco.2018.02.002

Aureli, S., Medei, R., Supino, E., & Travaglini, C. (2016, jan). Sustainability disclosure after a crisis. *International Journal of Social Ecology and Sustainable Development*, *7*(1), 35–49. doi: 10.4018/ijsesd.2016010102

Ballou, B., Heitger, D. L., Landes, C. H., & Adams, M. (2006). The future of corporate sustainability reporting. *Journal of Accountancy*, *202*(6), 70–74.

Bernard, S., Abdelgadir, S., & Belkhir, L. (2015, nov). Does GRI reporting impact environmental sustainability? an industry-specific analysis of CO2 emissions performance between reporting and non-reporting companies. *Journal of Sustainable Development*, *8*(9), 190. doi: 10.5539/jsd.v8n9p190

Bhimani, A., Silvola, H., & Sivabalan, P. (2016, mar). Voluntary corporate social responsibility reporting: A study of early and late reporter motivations and outcomes. *Journal of Management Accounting Research*, *28*(2), 77–101. doi: 10.2308/jmar-51440

Bloomfield, R. J. (2002). The 'incomplete revelation hypothesis' and financial reporting. *SSRN Electronic Journal*. doi: 10.2139/ssrn.312671

BMAS. (n.d). *Standards der CSR-berichterstattung.* https://www.csr-in-deutschland.de/DE/Unternehmen/CSR-Berichterstattung/Standards/standards.html. Retrieved 2022-05-19, from https://www.csr-in-deutschland.de/DE/CSR-Allgemein/CSR-in-der-Praxis/CSR-Berichterstattung/csr-berichterstattung.html

Caglio, A., Melloni, G., & Perego, P. (2020, oct). Informational content and assurance of textual disclosures: Evidence on integrated reporting. *European Accounting Review*, *29*(1), 55–83. doi: 10.1080/09638180.2019.1677486

Chauvey, J.-N., Giordano-Spring, S., Cho, C. H., & Patten, D. M. (2015, mar). The normativity and legitimacy of CSR disclosure: Evidence from france. *Journal of Business Ethics*, *130*(4), 789–803. doi: 10.1007/s10551-014-2114-y

Cho, C. H., & Patten, D. M. (2007, oct). The role of environmental disclosures as tools of legitimacy: A research note. *Accounting, Organizations and Society*, *32*(7-8), 639–647. doi: 10.1016/j.aos.2006.09.009

Cho, C. H., Roberts, R. W., & Patten, D. M. (2010, may). The language of US corporate environmental disclosure. *Accounting, Organizations and Society*, *35*(4), 431–443. doi: 10.1016/j.aos.2009.10.002

Christensen, H. B., Lee, E., Walker, M., & Zeng, C. (2015, jan). Incentives or standards: What determines accounting quality changes around IFRS adoption? *European Accounting Review*, *24*(1), 31–61. doi: 10.1080/09638180.2015.1009144

Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2008, may). Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. *Accounting, Organizations and Society*, *33*(4-5), 303–327. doi: 10.1016/j.aos.2007.05.003

Clarkson, P. M., Overell, M. B., & Chapple, L. (2011, mar). Environmental reporting and its relation to corporate environmental performance. *Abacus*, *47*(1), 27–60. doi: 10.1111/j.1467-6281.2011.00330.x

Clarkson, P. M., Ponn, J., Richardson, G. D., Rudzicz, F., Tsang, A., & Wang, J. (2020, mar). A textual analysis of US corporate social responsibility reports. *Abacus*, *56*(1), 3–34. doi: 10.1111/abac.12182

Cohen, J. R., Holder-Webb, L., & Zamora, V. L. (2015, jun). Nonfinancial information preferences of professional investors. *Behavioral Research in Accounting*, *27*(2), 127–153. doi: 10.2308/bria-51185

Corporate Finance Institute. (n.d). *What is the variance inflation factor (vif)?* https://corporatefinanceinstitute.com/resources/knowledge/other/variance-inflation-factor-vif/.

Davis, A. K., & Tama-Sweet, I. (2012, jan). Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&a∗. *Contemporary Accounting Research*, *29*(3), 804–837. doi: 10.1111/j.1911-3846.2011.01125.x

Dhaliwal, D. S., Li, O. Z., Tsang, A., & Yang, Y. G. (2011, jan). Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting. *The Accounting Review*, *86*(1), 59–100. doi: 10.2308/accr.00000005

Du, S., & Yu, K. (2020, apr). Do corporate social responsibility reports convey value relevant information? evidence from report readability and tone. *Journal of Business Ethics*, *172*(2), 253–274. doi: 10.1007/s10551-020-04496-3

Eccles, R. G., Krzus, M. P., & Solano, C. (2019). A comparative analysis of integrated reporting in ten countries. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3345590

Eikon. (2017). *Thomson reuters esg scores*. Thomson Reuters.

Elkington, J. (1998). Partnerships fromcannibals with forks: The triple bottom line of 21st-century business. *Environmental Quality Management*, *8*(1), 37–51. doi: 10.1002/tqem.3310080106

European Commission. (n.da). *Financial reporting.* https://ec.europa.eu/info/business-economy-euro/company-reporting-and-auditing/company-reporting/financial-reporting_en.

European Commission. (n.db). *Non-financial reporting.* https://ec.europa.eu/info/business-economy-euro/company-reporting-and-auditing/company-reporting/non-financial-reporting_en.

European Union. (2014). *Directive 2014/95/eu.* https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0095&from=DE.

Fiechter, P., Hitz, J.-M., & Lehmann, N. (2019). Real effects of disclosure regulation: Evidence from the european union's CSR directive. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3033883

Flesch, R. (n.d). *How to write plain english.* https://web.archive.org/web/20160712094308/http://www.mang.canterbury.ac.nz/writing_guide/writing/flesch.shtml.

Franco, G. D., Hope, O.-K., Vyas, D., & Zhou, Y. (2015, jul). Analyst report readability. *Contemporary Accounting Research*, *32*(1), 76–104. doi: 10.1111/1911-3846.12062

Global Reporting Initiative. (2016a). *Gri 101: Foundation.*

Global Reporting Initiative. (2016b). *Gri 102: General disclosures.*

Global Reporting Initiative. (2016c). *Gri 103: Management approach.*

Global Reporting Initiative. (2016d). *Gri 303: Water.*

Global Reporting Initiative. (n.da). *About gri.* https://www.globalreporting.org/information/about-gri/Pages/default.aspx.

Global Reporting Initiative. (n.db). *About sustainability reporting.* https://www.globalreporting.org/information/sustainability-reporting/Pages/default.aspx.

Global Reporting Initiative. (n.dc). *Gri and sustainability reporting.* https://www.globalreporting.org/information/sustainability-reporting/Pages/gri-standards.aspx.

Global Reporting Initiative. (n.dd). *Gri's history.* https://www.globalreporting.org/information/about-gri/gri-history/Pages/GRI's%20history.aspx.

Global Reporting Initiative. (n.de). *The gri standards: The global standards for sustainability reporting.* https://www.globalreporting.org/standards/media/2458/gri_standards_brochure.pdf.

Global Reporting Initiative. (n.df). *Sustainability reporting in the european union.* https://www.globalreporting.org/information/policy/Pages/EUpolicy.aspx.

GRI Secretariat. (2016). *Introducing the gri standards.* https://www.youtube.com/watch?time_continue=2&v=RDe8Kkh1BrU&feature=emb_title.

Huang, X., Nekrasov, A., & Teoh, S. H. (2012). Headline salience and over- and underreactions to earnings. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2101887

Hummel, K., & Rötzel, P. (2019, feb). Mandating the sustainability disclosure in annual reports—evidence from the united kingdom. *Schmalenbach*

Business Review, *71*(2), 205–247. doi: 10.1007/s41464-019-00069-8

Hummel, K., & Schlick, C. (2016, sep). The relationship between sustainability performance and sustainability disclosure – reconciling voluntary disclosure theory and legitimacy theory. *Journal of Accounting and Public Policy*, *35*(5), 455–476. doi: 10.1016/j.jaccpubpol.2016.06.001

Hussainey, K., & Walker, M. (2009, jan). The effects of voluntary disclosure and dividend propensity on prices leading earnings. *Accounting and Business Research*, *39*(1), 37–55. doi: 10.1080/00014788.2009.9663348

Ingram, R. W., & Frazier, K. B. (1980). Environmental performance and corporate disclosure. *Journal of Accounting Research*, *18*(2), 614. doi: 10.2307/2490597

International Integrated Reporting Council. (2013). *The international <ir> framework*. https://integratedreporting.org/wp-content/uploads/2015/03/13-12-08-THE-INTERNATIONAL-IR-FRAMEWORK-2-1.pdf.

INTOSAI. (2013). *Sustainability reporting: Concepts, frameworks and the role of supreme audit institutions*. https://www.wgea.org/media/2935/2013_wgea_sustainability-reporting.pdf.

Ioannou, I., & Serafeim, G. (2012, oct). What drives corporate social performance? the role of nation-level institutions. *Journal of International Business Studies*, *43*(9), 834–864. doi: 10.1057/jibs.2012.26

Ioannou, I., & Serafeim, G. (2017). *The consequences of mandatory corporate sustainability reporting*. Harvard Business School.

Kaplan, R. S., & Norton, D. P. (2008). Mastering the management system. *Harvard Business Review*, *86*(1), 62–77.

KPMG. (2017). *The road ahead–the kpmg survey of corporate responsibility reporting 2017*. https://home.kpmg/content/dam/kpmg/xx/pdf/2017/10/kpmg-survey-of-corporate-responsibility-reporting-2017.pdf.

Lang, M., & Stice-Lawrence, L. (2015, nov). Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics*, *60*(2-3), 110–135. doi: 10.1016/j.jacceco.2015.09.002

Latham, G. P., & Locke, E. A. (1979, sep). Goal setting—a motivational technique that works. *Organizational Dynamics*, *8*(2), 68–80. doi: 10.1016/0090-2616(79)90032-9

Leuz, C. (2010, jan). Different approaches to corporate reporting regulation: How jurisdictions differ and why. *Accounting and Business Research*, *40*(3), 229–256. doi: 10.1080/00014788.2010.9663398

Li, F. (2008, aug). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, *45*(2-3), 221–247. doi: 10.1016/j.jacceco.2008.02.003

Liew, W. T., Adhitya, A., & Srinivasan, R. (2014, apr). Sustainability trends in the process industries: A text mining-based analysis. *Computers in Industry*, *65*(3), 393–400. doi: 10.1016/j.compind.2014.01.004

Liu, M., Wu, K., Yang, R., & Yu, Y. (2020). Textual analysis for risk profiles from 10-k filings. *CPA Journal*, *90*(6), 36–41.

Locke, E. A., & Latham, G. P. (2002, sep). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705–717. doi: 10.1037/0003-066x.57.9.705

Loughran, T., & Mcdonald, B. (2011, jan). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, *66*(1), 35–65. doi: 10.1111/j.1540-6261.2010.01625.x

Loughran, T., & Mcdonald, B. (2016, jun). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187–1230. doi: 10.1111/1475-679x.12123

Lueg, K., Lueg, R., Andersen, K., & Dancianu, V. (2016, feb). Integrated reporting with CSR practices. *Corporate Communications: An International Journal*, *21*(1), 20–35. doi: 10.1108/ccij-08-2014-0053

Lukomnik, J., Kwon, S., & Welsh, H. (2018). State of sustainability and integrated reporting 2018. In (chap. IRRC Institute. Retrieved from IRRC Institute website: https://www.weinberg.udel.edu/IIRCiResearchDocuments/2018/11/2018-SP-500-Integrated-Reporting-FINAL-November-2018-1.pdf). Retrieved from IRRC Institute website: https: IRRC Institute.

Malmi, T., & Brown, D. A. (2008, dec). Management control systems as a package—opportunities, challenges and research directions. *Management Accounting Research*, *19*(4), 287–300. doi: 10.1016/j.mar.2008.09.003

Marx, B., & Mohammadali-Haji, A. (2014, apr). Emerging trends in reporting: An analysis of integrated reporting practices by south african top 40 listed companies. *Journal of Economic and Financial Sciences*, *7*(1), 231–250. doi: 10.4102/jef.v7i1.138

Melloni, G., Caglio, A., & Perego, P. (2017, may). Saying more with less? disclosure conciseness, completeness and balance in integrated reports. *Journal of Accounting and Public Policy*, *36*(3), 220–238. doi: 10.1016/j.jaccpubpol.2017.03.001

Miller, B. P. (2010, nov). The effects of reporting complexity on small and large investor trading. *The Accounting Review*, *85*(6), 2107–2143. doi: 10.2308/accr.00000001

Mittelbach-Hoermanseder, S., Hummel, K., & Rammerstorfer, M. (2019). Information content of corporate social responsibility disclosures in europe: An institutional perspective. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3365521

Muslu, V., Mutlu, S., Radhakrishnan, S., & Tsang, A. (2019, jan). Corporate social responsibility report narratives and analyst forecast accuracy. *Journal of Business Ethics*, *154*(4), 1119–1142. doi: 10.1007/s10551-016-3429-7

Muslu, V., Radhakrishnan, S., Subramanyam, K. R., & Lim, D. (2015, may). Forward-looking MD& a disclosures and the information environment. *Management Science*, *61*(5), 931–948. doi: 10.1287/mnsc.2014.1921

Nau, R. (n.d). *Regression diagnostics: Testing the assumptions of linear regression*. http://people.duke.edu/~rnau/testing.htm.

Nazari, J. A., Hrazdil, K., & Mahmoudian, F. (2017, aug). Assessing social and environmental performance through narrative complexity in CSR reports. *Journal of Contemporary Accounting & Economics*, *13*(2), 166–178. doi: 10.1016/j.jcae.2017.05.002

Neu, D., Warsame, H., & Pedwell, K. (1998, apr). Managing public impressions: Environmental disclosures in annual reports. *Accounting, Organizations and Society*, *23*(3), 265–282. doi: 10.1016/s0361-3682(97)00008-1

Patten, D. M. (2002, nov). The relation between environmental performance and environmental disclosure: a research note. *Accounting, Organizations and Society*, *27*(8), 763–773. doi: 10.1016/s0361-3682(02)00028-4

Pencle, N., & Mălăescu, I. (2016, sep). What's in the words? development and validation of a multidimensional dictionary for CSR and application using prospectuses. *Journal of Emerging Technologies in Accounting*, *13*(2), 109–127. doi: 10.2308/jeta-51615

Pizzetti, M., Gatti, L., & Seele, P. (2019, dec). Firms talk, suppliers walk: Analyzing the locus of greenwashing in the blame game and introducing 'vicarious greenwashing'. *Journal of Business Ethics*, *170*(1), 21–38. doi: 10.1007/s10551-019-04406-2

Purplemath. (n.d). *Interquartile ranges & outliers*. https://www.purplemath.com/modules/boxwhisk3.htm.

Purvis, B., Mao, Y., & Robinson, D. (2019, sep). Three pillars of sustainability: in search of conceptual origins. *Sustainability Science*, *14*(3), 681–695. doi: 10.1007/s11625-018-0627-5

PwC. (2014). *Sustainability goes mainstream: Insights into investor views*. https://www.pwc.com/us/en/pwc-investor-resource-institute/publications/assets/pwc-sustainability-goes-mainstream-investor-views.pdf.

SEC. (1998). *A plaing english handbook: How to create clear sec disclosure documents*. U.S. Securities and Exchange Commission, Washinton, DC.

Statistics Solutions. (2013). *Homoscedasticity*. https://www.statisticssolutions.com/homoscedasticity/.

Statistics Solutions. (n.d). *Assumptions of linear regression*. https://www.statisticssolutions.com/assumptions-of-linear-regression/.

Stent, W., Bradbury, M., & Hooks, J. (2013, sep). What firms' discretionary narrative disclosures reveal about the adoption of international financial reporting standards. *Australian Accounting Review*, *23*(3), 252–263. doi: 10.1111/auar.12002

STOXX. (2020). *Changes in composition of stoxx benchmark indices*. https://www.stoxx.com/press-releases-details?articleId=1846984593.

United Nations. (1987). *Report of the world commission on environment and*

development: Our common future.

Wakefield, J. (2016). *Apple, samsung and sony face child labour claims [news website]*. https://www.bbc.com/news/technology-35311456.

Wang, Z., Hsieh, T.-S., & Sarkis, J. (2018, jul). CSR performance and the readability of CSR reports: Too good to be true? *Corporate Social Responsibility and Environmental Management*, *25*(1), 66–79. doi: 10.1002/csr.1440

Wiseman, J. (1982, jan). An evaluation of environmental disclosures made in corporate annual reports. *Accounting, Organizations and Society*, *7*(1), 53–63. doi: 10.1016/0361-3682(82)90025-3

Yang, F., Dolar, B., & Mo, L. (2018, mar). Textual analysis of corporate annual disclosures: A comparison between bankrupt and non-bankrupt companies. *Journal of Emerging Technologies in Accounting*, *15*(1), 45–55. doi: 10.2308/jeta-52085

Yusof, S. M. (2018). *Social environmental disclosure between gri-sustainability reporting and iirc – integrated reporting among european companies.*

Junior Management Science 7(3) (2022) 604-642

Junior Management Science

journal homepage: www.jums.academy

Effects of the European CSR Directive (2014/95/EU) on the Credibility of Sustainability Reporting – An Empirical Study of German Listed Companies

Auswirkung der europäischen CSR-Richtlinie (2014/95/EU) auf die Glaubwürdigkeit der Nachhaltigkeitsberichterstattung – Eine empirische Untersuchung deutscher börsennotierter Unternehmen

Tim Scheufen

Rheinisch-Westfälische Technische Hochschule Aachen

Abstract

The European Union has implemented the CSR Directive 2014/95/EU with the aim to improve sustainability reporting in order to increase the trust of different stakeholders. There is evidence in the literature that stakeholders distrust the sustainability reporting of companies. The lack of credibility may lead to a loss legitimacy or acceptance in society. Accordingly, the master thesis raises the research question whether the implemented CSR directive has increased credibility in the sustainability reporting. To answer the research question the credibility of the sustainability reporting is quantified before and after the implementation of the CSR directive in a German sample (MDAX). Credibility is measured by a credibility index which consists of the following three dimensions: (1) Truth, (2) sincerity and (3) appropriateness & understandability. The results show a statistically significant increase in the dimensions of truth and sincerity after the introduction of the CSR directive. Therefore, the master thesis delivers empirical evidence for the effectiveness of the implemented CSR directive in the context of credibility.

Zusammenfassung

Die Europäische Union hat die CSR-Richtlinie 2014/95/EU auf den Weg gebracht, um die Nachhaltigkeitsberichterstattung zu verbessern und damit das Vertrauen der Stakeholder zu stärken. Die bestehende Literatur weist darauf hin, dass zahlreiche Stakeholder der Nachhaltigkeitsberichterstattung von Unternehmen mit Misstrauen gegenüberstehen. Die fehlende Glaubwürdigkeit führt dazu, dass das Unternehmen seine Legitimität bzw. Akzeptanz in der Gesellschaft verliert. Aus diesem Grund untersucht die Masterarbeit, ob die implementierte CSR-Richtlinie die Glaubwürdigkeit in der Nachhaltigkeitsberichterstattung erhöht hat. Diese Fragestellung wird anhand einer deutschen Stichprobe untersucht (MDAX). Hierfür wird die Glaubwürdigkeit der Nachhaltigkeitsberichterstattung vor und nach der Einführung der CSR-Richtlinie quantifiziert. Die Messung der Glaubwürdigkeit erfolgt mit Hilfe eines Glaubwürdigkeitsindex, welcher aus den folgenden drei Dimensionen besteht: (1) Wahrheit, (2) Wahrhaftigkeit sowie (3) Angemessenheit & Verständlichkeit. Die Ergebnisse zeigen einen statistisch signifikanten Anstieg der Dimensionen Wahrheit und Wahrhaftigkeit nach der Einführung der CSR-Richtlinie, was für die Wirksamkeit der durchgeführten CSR-Richtlinie hinsichtlich der Glaubwürdigkeit spricht.

Keywords: CSR-Richtlinie 2014/95/EU; Glaubwürdigkeit; Nachhaltigkeitsberichterstattung; MDAX.

1. Einleitung

Das Bedürfnis der Stakeholder[1], über nachhaltigkeitsbezogene Inhalte informiert werden zu wollen, ist in den letzten Jahren sukzessive gewachsen (Daub, 2010, S. 30). Folglich wird die traditionelle Finanzberichterstattung oftmals durch eine gesonderte Nachhaltigkeitsberichterstattung (NBE) ergänzt (Frias-Aceituno, Rodríguez-Ariza & García-Sánchez, 2013, S. 45). Obwohl stetig mehr nichtfinanzielle Informationen (NFI) von Unternehmen publiziert werden, konnte das Misstrauen der Stakeholder in die NBE nicht beseitigt werden. Wider der Erwartung stieg das Misstrauen sogar

[1] Grundsätzlich wird im Rahmen der Masterarbeit bei Personenbezeichnungen und personenbezogenen Hauptwörtern die männliche Form verwendet. Dies erfolgt aus dem Grund der besseren Lesbarkeit und impliziert keinerlei Wertung. Im Sinne der Gleichbehandlung werden grundsätzlich alle Geschlechter angesprochen.

DOI: https://doi.org/10.5282/jums/v7i3pp604-642

(Spelthahn, Fuchs & Demele, 2009, S. 61; Waddock & Goggins, 2011, S. 25). Die Skepsis der Stakeholder artikuliert sich bspw. wie folgt: „Corporate social responsibility? Isn't that an oxymoron?", „Corporate social responsility, that's just windowdressing" oder "That's just greenwashing." (Waddock & Goggins, 2011, S. 27) Im Werbefilm des Energiekonzerns RWE streift im Jahre 2009 der grüne Riese durch die idyllische Natur und „pflanzt" fröhlich Wind- sowie Wasserkraftanlagen. Zu dieser Zeit lag der Anteil erneuerbarer Energien im Energiemix von RWE bei 2 % (Bentele & Nothhaft, 2011, S. 45-46). Durch solche unglaubwürdigen Werbebotschaften fühlen sich Stakeholder darin bestärkt, dass Unternehmen die Kommunikation über Nachhaltigkeit nur zum Zwecke von wirtschaftlichen Interessen nutzen (Spelthahn et al., 2009, S. 61-62). Dabei sollte es im Interesse des Unternehmens sein, als glaubwürdig von seinen Stakeholdern wahrgenommen zu werden. Eine als glaubwürdig wahrgenommene NBE führt dazu, dass einem Unternehmen Legitimität zugesprochen wird (Lock, 2016, S. 416-417). Ein gesellschaftlich legitimiertes Unternehmen zu sein, bietet Vorteile, wie z.B. geringere Finanzierungskosten, bessere Bewerberpools, den Aufbau von Reputation, u.v.m. (M. Fifka, 2014, S. 11-12; Sandhu, 2012, S. 167). In Folge mangelnder Legitimation durch die Stakeholder kann ein Unternehmen schlimmstenfalls gar boykottiert werden, in die öffentliche Kritik geraten oder rechtliche Konsequenzen auslösen (Spelthahn et al., 2009, S. 62). Das Hervorbringen einer glaubwürdigen NBE kann als zentrales Instrumentarium angesehen werden, gesellschaftliche Akzeptanz bzw. Legitimität zu generieren (Lock, 2016, S. 416-417).

Die zunehmende Relevanz der NFI für Stakeholder führte in den letzten Jahren dazu, dass weltweit zunehmend regulatorische Maßnahmen von politischer Seite getroffen wurden, um die Transparenz publizierter Inhalte im Kontext der Nachhaltigkeit zu erhöhen (Carrots & Sticks, 2016, S. 9; Gulenko, 2018, S. 3). Ebenso führte die Europäische Union (EU) die CSR-Richtlinie 2014/95/EU (CSR-RL) ein, wonach die europäischen Mitgliedsstaaten verpflichtet wurden, strengere Vorschriften in der NBE zu berücksichtigen (EU, 2014b, S. 1-9). Die Europäische Kommission verspricht sich u.a. Folgendes von der Einführung der CSR-RL:

> „Each individual company disclosing transparent information on social and environmental matters will realise significant benefits over time, including [...] better relations with consumers and stakeholders. Investors and lenders will benefit from a more informed and efficient investment decision process." (EU, 2014a)

Faktisch mangelt es der EU an empirischer Evidenz aus der Forschung, welche untersucht, ob das angestrebte Ziel, das Verhältnis zwischen Unternehmen und Stakeholdern zu verbessern, erreicht werden konnte (Maas & Sampers, 2020, S. 266). Des Weiteren existieren in der Literatur ambivalente Ergebnisse hinsichtlich der Frage, ob eine verpflichtende NBE die Glaubwürdigkeit erhöht (Gulenko, 2018, S. 11). Aus diesem Grund wirft die Masterarbeit folgende Forschungsfrage auf: Welche Auswirkungen hat die CSR-RL auf die Glaubwürdigkeit der NBE?

Um das Forschungsvorhaben umzusetzen, wurde folgende inhaltliche Strukturierung vorgenommen: Zunächst erfolgt eine inhaltliche Abgrenzung des Terminus Glaubwürdigkeit. Nachdem über das Drei-Säulen-Modell typische Inhalte der NBE herausgearbeitet werden, wird die Bedeutung der Legitimität für ein Unternehmen beschrieben. Hierbei wird auf die Umwelt eines Unternehmens, das Vier-Stufen-Pyramiden-Modell sowie die Legitimitätstheorie eingegangen. Danach erfolgt eine Erläuterung jener Indikatoren, die gemäß der aktuellen Forschung für eine glaubwürdige NBE stehen. Anschließend wird jeweils dem regulatorischen Hintergrund, dem aktuellen Forschungsstand sowie der Hypothesenherleitung ein Kapitel gewidmet. Aus der Theorie und der Empirie leitet sich die Hypothese ab, dass die CSR-RL zu einer glaubwürdigeren NBE führt. Der Forschungsgegenstand wird mit Hilfe einer Inhaltsanalyse als Forschungsmethodik untersucht. Um die Glaubwürdigkeit der NBE zu messen, nutzt die Masterarbeit einen bereits bestehenden Glaubwürdigkeitsindex. Der verwendete Glaubwürdigkeitsindex besteht aus den drei Dimensionen: Wahrheit, Wahrhaftigkeit und Angemessenheit & Verständlichkeit (Mazzotta, Bronzetti & Veltri, 2020, S. 1907). Diese leiten sich aus der Theorie des kommunikativen Handelns von Jürgen Habermas ab (Habermas, 2002). Bei der betrachteten Stichprobe handelt es sich um die Unternehmen des MDAX. Die Implementierung der CSR-RL durch den deutschen Gesetzgeber erfolgte in dem sog. CSR-Richtlinien-Umsetzungsgesetz (CSR-RUG). Die Anwendung der Berichtspflicht trat für deutsche Unternehmen erstmalig für Geschäftsjahre, die nach dem 31.12.2016 beginnen, ein (DNK, 2018, S. 6). Dies ermöglicht eine Ex-Post-Evaluation der Glaubwürdigkeit in der NBE vor und nach der Einführung des Gesetzes. In der Literatur wurden Einflussgrößen ermittelt, die einen Einfluss auf die Glaubwürdigkeit der NBE ausüben können (Ali, Frynas & Mahmood, 2017, S. 276-287; Dienes, Sassen & Fischer, 2016, S. 167). Für diese möglichen Störeinflüsse wird mittels einer multiplen linearen Regressionsanalyse kontrolliert. Die Veränderung der Glaubwürdigkeit wird mit Hilfe von deskriptiver und induktiver Statistik ausgewertet. Ein Fazit fasst die Kernergebnisse am Ende nochmals zusammen.

2. Glaubwürdigkeit in der Nachhaltigkeitsberichterstattung

Nachdem die begrifflichen Grundlagen, wie z.B. Glaubwürdigkeit oder NFI, erläutert werden, wird danach auf die Beziehung eines Unternehmens mit seiner Umwelt ausführlich eingegangen. Die Rolle der Legitimität wird hierbei in den Vordergrund gestellt. Nachdem eine glaubwürdige NBE als Voraussetzung für Legitimität identifiziert wird, beschäftigt sich das letzte Unterkapitel mit Indikatoren für eine glaubwürdige NBE, die bislang in der Forschung festgestellt werden konnten.

2.1. Definitorische Abgrenzung der Glaubwürdigkeit

Es gibt aus kommunikationswissenschaftlicher Sicht immer einen Kommunikator, der eine Botschaft an einen oder mehrere Informationsempfänger (Rezipienten) übermittelt (Spelthahn et al., 2009, S. 62). Die Frage, ob eine Botschaft als glaubwürdig einzustufen ist oder nicht, stellt sich immer dann, wenn eine gewisse Unsicherheit über den Wahrheitsgehalt einer Botschaft besteht. Eine kommunikationszentrierte Betrachtungsweise impliziert, dass eine Botschaft schon dann als glaubwürdig einzustufen ist, wenn der Kommunikator die Botschaft als zutreffend erachtet und keine Täuschungsabsicht vorliegt (Köhnken, 1990, S. 4). Die rezipientenzentrierte Perspektive geht hingegen davon aus, dass Glaubwürdigkeit eine Eigenschaft ist, die ausschließlich vom Rezipienten der Kommunikationsquelle zugesprochen werden kann (Eisend, 2003, S. 39). Im Rahmen der Masterarbeit ist die rezipientenzentrierte Betrachtungsweise relevant, da die Eigenschaft einer glaubwürdigen Kommunikation nur durch die Stakeholder als Rezipienten bewertet werden kann. Die Zu- oder Absprache von Glaubwürdigkeit ist das Resultat dieser Evaluation.

In der Alltagssprache werden die Begriffe Glaubwürdigkeit und Vertrauen meist synonym verwendet. Dabei weisen beide Begriffe einen unterschiedlichen Zeit- und Objektbezug auf. Vertrauen ist auf die Zukunft und eher auf Objekte gerichtet. Zum Beispiel wird einem Flugzeug Vertrauen geschenkt, in dem man hofft, sicher zu reisen. Glaubwürdigkeit ist hingegen gegenwarts- bzw. vergangenheitsorientiert und wird Subjekten zu- bzw. abgesprochen (Spelthahn et al., 2009, S. 63). Im Kontext der NBE können die Begriffe Glaubwürdigkeit und Vertrauen wie folgt abgegrenzt werden: Bei der Evaluation der NBE durch die Stakeholder ist die Glaubwürdigkeit betroffen, da es sich um eine Bewertung von vergangenheitsorientierten Inhalten handelt. Hingegen vertrauen die Stakeholder dem Unternehmen dahingehend, dass bspw. definierte Nachhaltigkeitsziele zukünftig tatsächlich umgesetzt werden.

2.2. Das Instrumentarium der Nachhaltigkeitsberichterstattung

1987 entwickelte die Sachverständigenkommission der Vereinten Nationen unter der Leitung der norwegischen Ministerpräsidentin Gro Harlem Brundtland eine Definition für eine nachhaltige Entwicklung: Eine nachhaltige Entwicklung „[...] meets the needs of the present without compromising the ability of the future generations to meet their own needs.“ (Brundtland, 1987, S. 15) Die Definition bezieht sich vor allem auf die intra- und die intergenerationale Gerechtigkeit. Dieses normative Leitbild schreibt vor, dass bestimmte Menschen nicht auf Kosten anderer Menschen in anderen Regionen oder künftiger Generationen leben sollen (Kropp, 2018, S. 5). Bis heute entwickelt sich das Verständnis, was Nachhaltigkeit in verschiedenen Kontexten bedeutet, kontinuierlich weiter (Moore, Mascarenhas, Bain & Straus, 2017, S. 5). Dennoch kann die Brundtland-Definition als Minimalkonsens beurteilt werden, wie das Konzept einer nachhaltigen Entwicklung zu verstehen ist. Eine nachhaltige Entwicklung entspricht einem dynamischen Prozess, wobei der Begriff der Nachhaltigkeit als angestrebtes Ziel anzusehen ist (Kropp, 2018, S. 6). Im Drei-Säulen-Modell (Abb. 1) wird die Nachhaltigkeit als „Dach“ dargestellt, welches auf drei „Säulen“ steht. Die Säulen Soziales, Wirtschaft und Umwelt stehen gleichberechtigt nebeneinander (Kropp, 2018, S. 11). Demnach trägt bspw. die Gewinnmaximierung nur zur Nachhaltigkeit bei, wenn sie sowohl umwelt- als auch sozialverträglich ist (Elkington, 1997, S. 49). Grundsätzlich tragen alle Individuen und Organisationen, wie z.B. Unternehmen, Staaten, private Haushalte etc., die Verantwortung, das Ziel der Nachhaltigkeit voranzutreiben (Kropp, 2018, S. 17-22). Der privatwirtschaftliche Sektor und insbesondere multinationale Konzerne werden als wesentliche Akteure angesehen, die eine besonders große Wirkung erzielen können. Dies wird bspw. mit der Finanzstärke und dem vorhandenen Humankapital begründet (Berrone et al., 2019, S. 16).

Während die finanzielle Publizitätspflicht bereits ihren Ursprung im 13. Jahrhundert in europäischen Städten hatte (Merkt, 2001, S. 31-22), handelt es sich bei der NBE um eine Praxis, welche in Westeuropa erst in den 1970er aufkam (M. Fifka, 2014, S. 3). Der gesellschaftliche Druck über ökologische und soziale Themen zu berichten, ist seit dem ständig gewachsen, sodass die Unternehmen die NBE als Kommunikationsmedium verwenden, um über nachhaltigkeitsbezogene Themen zu berichten (Daub, 2010, S. 30).

Die NBE befasst sich thematisch mit den drei Säulen Soziales, Umwelt sowie Wirtschaft (M. Fifka, 2014, S. 4). In Tabelle 1 sind beispielhaft Themen dargestellt, auf die sich die NBE beziehen kann (GRI, 2016). Darüber hinaus zählen auch Informationen über die Unternehmensführung (Corporate Governance) sowie das Ethik-Management zu möglichen Themenbereichen der NBE. Die Unternehmensführung beschäftigt sich mit den Leitungs- und Kontrollstrukturen, während sich das Ethik-Management mit der Erstellung eines Code of Conducts befasst (M. Fifka, 2014, S. 5). Im Rahmen der Masterarbeit beschränkt sich die NBE analog zu anderen wissenschaftlichen Artikeln auf die Säulen Soziales und Umwelt sowie die Unternehmensführung. Der Begriff der nichtfinanziellen Informationen (NFI) umfasst ausschließlich diese Themenbereiche (z.B. Dumay, Frost & Beck, 2015, S. 2).

Die NBE kann entweder in den Lagebericht integriert werden oder aber das Unternehmen entscheidet sich dafür, über NFI in einem gesonderten Bericht zu informieren. Sofern ein gesonderter Bericht erstellt wird, kann es zu Schnittmengen zwischen dem Lagebericht und dem gesonderten Bericht kommen. Demnach berichtet ein Unternehmen u.U. sowohl über Emissionen, Energieverbrauch, Fluktuation der Mitarbeiter, etc. im Lagebericht als auch in dem gesonderten Bericht (Lackmann, 2010, S. 32). Für einen gesonderten Bericht haben sich in der Praxis verschiedene Bezeichnungen entwickelt. Die Bezeichnungen CSR Report bzw. Nachhaltigkeitsbericht (Sustainability Report) haben sich jedoch weitestgehend durchgesetzt (M. Fifka, 2014, S. 3-4). In der Empirie zeigte sich kein signifikanter Unterschied zwischen den In-

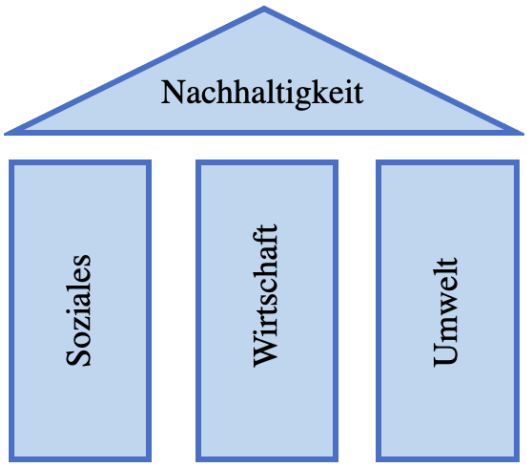


Abbildung 1: Drei-Säulen-Modell (in Anlehnung an Pufé, 2017, S. 110)

Tabelle 1: Mögliche Themen der NBE (eigene Darstellung)

Soziales	Wirtschaft	Umwelt
Einhaltung der Menschenrechte	Unmittelbar erzeugter und ausgeschütteter wirtschaftlicher Wert	Emissionsreduktion und Klimaschutz
Karriere- und Weiterbildungsmöglichkeiten für die Belegschaft	Finanzielle Unterstützung durch die öffentliche Hand	Ressourcen- und Energieeffizienz
Diversität und die Gleichstellung von Mann und Frau	Rechtsverfahren aufgrund von wettbewerbswidrigem Verhalten, Kartell- und Monopolbildung	Abfallmanagement und Recycling
Sicherstellung sozialer Standards in der Lieferkette	Maßnahmen zur Korruptionsbekämpfung	Sicherstellung einer ökoloschen Lieferkette

halten mit den verschiedenen Namensgebungen (M. Fifka, 2014, S. 3; M. S. Fifka & Drabble, 2012, S. 468-470).

2.3. Die Bedeutung der Legitimität für ein Unternehmen

Nach der Erläuterung, was ein Stakeholder überhaupt ist und wie ein typisches Stakeholder Netzwerk eines Unternehmens aussieht, wird im Anschluss auf die Legitimitätstheorie sowie auf den Zusammenhang zwischen Glaubwürdigkeit und Legitimität in der NBE eingegangen.

2.3.1. Interne und externe Stakeholder

Ein Stakeholder ist definiert als „any individual or group affect or is affected by the actions, decision, policies ‚practices, or goals of the organization." (Carroll & Buchholtz, 2006, S. 66) In Abbildung 2 sind ausgewählte Anspruchsgruppen dargestellt, die ein Unternehmen haben kann.

Alle Stakeholder bzw. Anspruchsgruppen haben gemeinsam, dass sie ein berechtigtes Interesse ("Stake") an einer Organisation besitzen (Carroll & Buchholtz, 2006, S. 65). Zu unterscheiden sind sog. primäre (interne) und sekundäre (externe) Stakeholder. Primäre Stakeholder (z.B. Aktionäre, Arbeitnehmerschaft, Lieferanten) haben im Gegensatz zum sekundären Stakeholder (z.B. NGOs, Medien, Konkurrenzunternehmen) einen unmittelbaren Einfluss auf das Unternehmen und bestimmen den Erfolg maßgeblich. Ein Unternehmen operiert folglich in einer komplexen Umwelt, die aus diversen Anspruchsgruppen bestehen kann (Carroll & Buchholtz, 2006, S. 68). Zudem wird in der Literatur zwischen drei verschiedenen „Stakes" unterschieden, die im Folgenden im Kontext eines Unternehmens erläutert werden: Das erste „Stake" bezieht sich auf rechtlich durchsetzbare Ansprüche. Zum Beispiel gibt es Eigentümer, die Anteile an einem Unternehmen halten und die damit verbundenen Rechte besitzen (z.B. Ausschüttung des Residualgewinns in Form von Dividenden, Stimmrecht, etc.). Ferner kann ein anderes im Gesetz verankertes Recht bestehen, einen bestimmten Anspruch, wie z.B. den gesetzlichen Mindestlohn, durchzusetzen (Carroll & Buchholtz, 2006, S. 65). Ebenso existiert ein „Stake" in Form eines moralischen Rechts, sodass bspw. distributionale- und prozedurale Gerechtigkeit in einem Unternehmen eingefordert werden kann (Carroll & Buchholtz, 2006, S. 65). Das letztes „Stake" beschreibt ein berechtigtes Interesse, welches ein Stakeholder haben kann, da er in bestimmter Art und Weise durch die Entscheidungen eines Unternehmens tangiert wird. Dies ist z.B. der Fall, wenn ein Unternehmen schädliche Stoffe emittiert, welche für die umliegende Bevölkerung gesundheitsgefährdend sind (Carroll & Buchholtz, 2006, S. 65).

Abbildung 2: Mögliche Stakeholder eines Unternehmens (eigene Darstellung)

2.3.2. Die Vier-Stufen-Pyramide

Der populäre US-Ökonom und Wirtschaftsnobelpreisträger Milton Friedman äußerte sich 1962 wie folgt zur unternehmerischen Verantwortung: „[In a free society] ... there is one and only one social responsibility of business – to use its resources and engage in activities designed to increase its profits." (Friedman, 2007, S. 178) Dieses Zitat verdeutlicht, dass die Eigentümer einer Unternehmung und die Gewinnmaximierung nach Friedman im Vordergrund stehen (Ferrero, Michael Hoffman & McNulty, 2014, S. 39-43). Demgegenüber existiert die Idee der unternehmerischen Verantwortung, welche im Englischen unter dem Begriff Corporate Social Responsibility (CSR) bekannt ist. Demnach trägt ein Unternehmen sehr wohl eine Verantwortung für die Gesellschaft (Carroll, Lipartito, Post & Werhane, 2012, S. 31-376). In Abbildung 3 ist die Vier-Stufen-Pyramide dargestellt (Carroll, 1979, 1991):

Nach Carroll (1979, S. 500) wird CSR wie folgt definiert: „The social responsibility encompasses the economic, legal, ethical, and discretionary [philanthropic] expectations that society has of organizations at a given point in time." Das Fundament stellt die ökonomische Verantwortung dar, das den künftigen Fortbestand eines Unternehmens sichert (Carroll, 2016, S. 3). Die rechtliche Verantwortung fordert die Einhaltung von Gesetzen, welche eine kodifizierte Erwartungshaltung der Gesellschaft darstellt (Carroll, 2016, S. 3-4). Neben expliziten Gesetzen erwartet die Gesellschaft jedoch ebenfalls die Beachtung impliziter ethischer Normen, die nicht im Gesetz verankert sind, allerdings trotzdem berücksichtigt werden sollten (Carroll, 2016, S. 4). Um die rechtliche und ethische Verantwortung zu differenzieren, dient folgendes Beispiel: Für multinationale Konzerne ist es legal, Steuerumgehungsstrategien zu nutzen. Allerdings stellt sich die Frage, ob es als ethisch vertretbar anzusehen ist, dass die relevanten Länder, in denen der Absatz erzielt wird, nicht von den Steuern in Form von Gemeinwohlinvestitionen profitieren sollten (Walden & Depping, 2015, S. 218-219). Die philanthropische Verantwortung inkludiert alle Aktivitäten, die von den vorausgegangenen Verantwortungsbereichen nicht einbezogen werden (Carroll, 2016, S. 4). Dieser Verantwortungsbereich ist mit dem Corporate Citizenship („The Art of Giving Back to the Community") gleichzusetzen. Hier sind vor allem das Corporate Giving (z.B. Geld- oder Sachspenden), Corporate Foundations (z.B. Unternehmensstiftungen) oder auch das Corporate Volunteering (z.B. Freistellung von Mitarbeitenden für ehrenamtliche Tätigkeiten) zu nennen (Loew, Ankele, Braun & Clausen, 2004, S. 53). Die Darstellung in Form einer Pyramide verdeutlicht, dass ein Unternehmen zu aller erst wirtschaftlich rentabel und gesetzeskonform agieren muss, bevor es ethische oder philanthropische Aspekte implementieren kann. Der ethische Verantwortungsbereich wird von der Gesellschaft erwartet und der philanthropische Verantwortungsbereich wird sukzessive für Stakeholder wichtiger (Carroll, 2016, S. 4).

2.3.3. Legitimitätstheorie

Ein Unternehmen besitzt nicht naturgegeben und uneingeschränkt das Recht wirtschaftlich aktiv zu sein. Viel mehr benötigt es von seiner Umwelt bzw. seinen Stakeholdern Legitimation bzw. die sog. „License to Operate" (Demuijnck & Fasterling, 2016, S. 677-679). Folgende Definition fasst zusammen, was der Begriff Legitimität bedeutet:

> „Legitimität ist eine verallgemeinerte Wahrnehmung oder Annahme, dass die Handlungen eines sozialen Gebildes innerhalb eines sozial konstruierten Systems von Normen, Werten, Glaubensannahmen und Definitionen als erstrebenswert,

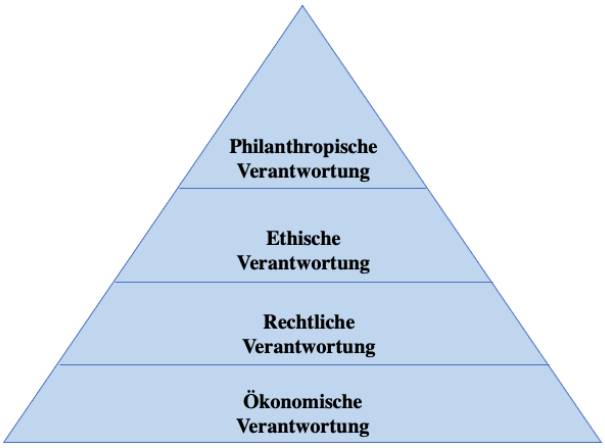


Abbildung 3: Vier-Stufen-Pyramide (in Anlehnung an Carroll, 2016, S. 5)

> zweckmäßig oder angemessen gelten.“ (Sandhu, 2012, S. 165)

Ein Unternehmen kann als legitim oder illegitim von seiner Umwelt angesehen werden. Die Zuschreibung von Legitimität erfolgt nicht durch einzelne Individuen, sondern dies geschieht nur durch soziale Gruppen mit gleichen Normen, Werten, Glaubensannahmen und Definitionen (Berger & Luckmann, 2004, S. 62). Dabei resultiert die Legitimität aus der umfassenden Historie bzw. der Erfahrung mit dem Unternehmen und kurzfristige Ereignisse führen nicht unmittelbar zur Absprache von Legitimität (Sandhu, 2012, S. 166). Obgleich ein Unternehmen seine Legitimität meist nicht durch einzelne Vorfälle vollständig einbüßt, kann sie jedoch z.B. durch Skandale oder moralisch fragwürdige Handlungsweisen angegriffen werden (Demuijnck & Fasterling, 2016, S. 678; Suchman, 1995, S. 574). Gemäß der Literatur sind die pragmatische-, die moralische- sowie die kognitive Legitimität voneinander zu unterscheiden:

Ein Unternehmen sollte aus rein strategischen Erwägungen nach Legitimität streben, da der Zugang zu Ressourcen, wie z.B. Kapital, Personal, etc., erleichtert wird (M. Fifka, 2014, S. 11-12; Sandhu, 2012, S. 167). Zudem kann ein Unternehmen, das nicht den Erwartungen seiner Umwelt entspricht, als unzuverlässig, irrational oder unnötig stigmatisiert werden und damit einen Reputationsverlust erleiden (Sandhu, 2012, S. 167; Suchman, 1995, S. 575). Diese Form der Legitimität wird auch als pragmatische Legitimität bezeichnet (Lock, 2016, S. 419). Das kalkulierbare Eigeninteresse eines Unternehmens steht hierbei im Vordergrund (Sandhu, 2012, S. 169; Suchman, 1995, S. 578).

Moralische Legitimität liegt dann vor, wenn der Abgleich zwischen dem, was das Unternehmen tut und dem, was die Umwelt verlangt bzw. als richtig ansieht, positiv für das Unternehmen ausfällt (Lock, 2016, S. 419). Konkret ist moralische Legitimität definiert als „[...] die wertbezogene, positive Beurteilung einer Organisation [...]“ und diese „[...] liegt vor, wenn eine Organisation die wertbasierten Anforderungen ihrer Umwelt erfüllt.“ (Sandhu, 2012, S. 171)

Die dritte Form der Legitimität ist die sogenannte kognitive Legitimität. Zum einen trägt ein einfach zu verstehendes Geschäftsmodell (z.B. Apotheke, Bäckerei) dazu bei, dass eine Unternehmung als legitim angesehen wird. Außerdem gibt es Unternehmen, die per se eine hohe Legitimität genießen (z.B. Krankenhäuser, Anwaltskanzleien), da die Gesellschaft ohne sie nicht funktionieren würde (Sandhu, 2012, S. 167, 173; Suchman, 1995, S. 582).

In Abbildung 4 ist nochmals das Zusammenspiel von Unternehmen mit dessen Umwelt bzw. Stakeholder dargestellt. Die Literatur geht davon aus, dass ein impliziter sozialer Vertrag zwischen einem Unternehmen und der Gesellschaft bzw. verschiedenen Stakeholdergruppen besteht (Carroll & Buchholtz, 2006, S. 17). Im gemeinsamen Austausch muss das Unternehmen ständig beweisen, dass es mit den normativen Erwartungen der Stakeholder in Einklang steht, da die Legitimität von der Erfüllung dieses (impliziten) Vertrages abhängig ist (Cho, Laine, Roberts & Rodrigue, 2015, S. 80). Dieser Aspekt bezieht sich wie oben erläutert auf die moralische Legitimität. Die Einhaltung von Gesetzen ist hingegen zwingend für ein Unternehmen. Die Vermeidung einer Strafzahlung oder negativer öffentlicher Kritik durch einen Gesetzesverstoß beruht auf dem Eigeninteresse des Unternehmens und ist demnach der pragmatischen Legitimität zuzuordnen. Kognitiv legitimierte Unternehmen werden gesellschaftlich nicht mehr hinterfragt und sind entsprechend weniger von der pragmatischen- bzw. moralischen Legitimität abhängig.

2.4. Glaubwürdigkeit als Voraussetzung für Legitimität

Das Konzept der Legitimität ist eng verwoben mit Glaubwürdigkeit. Eine als glaubwürdig wahrgenommene NBE führt dazu, dass einem Unternehmen Legitimität zugesprochen wird. Vice versa führt eine unglaubwürdige NBE zu einer Minderung der Legitimität (Lock, 2016, S. 416-417). Damit ist eine glaubwürdige Kommunikation integraler Bestandteil für eine langlebige und vertrauensvolle Beziehung mit seinen Stakeholdern (Lock, 2016, S. 414). Moralische Legitimität kann in einem deliberativen Austausch, z.B. mittels der NBE, zwischen einem Unternehmen und seiner Umwelt erzielt werden (Seele & Lock, 2015, S. 403).

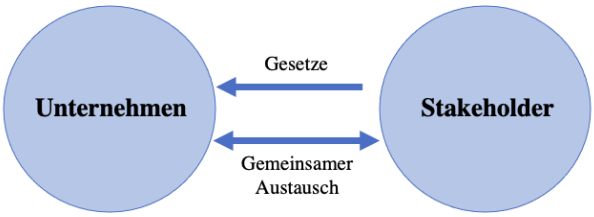


Abbildung 4: Unternehmen und Stakeholder im ständigen Austausch (in Anlehnung an Carroll & Buchholtz, 2006, S. 18)

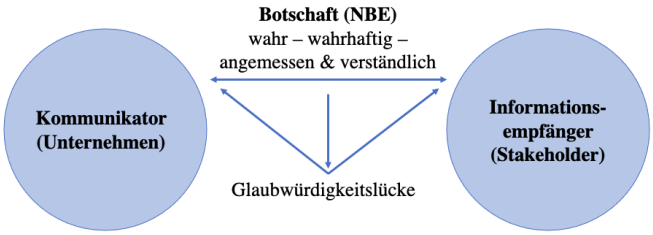


Abbildung 5: Kommunikationsmodell - Legitimitätsbasierter Glaubwürdigkeitsansatz (in Anlehnung an Seele & Lock, 2015, S. 404)

In Abbildung 5 ist ein kommunikationswissenschaftliches Modell abgebildet. Der Kommunikator ist das Unternehmen. Die Botschaft ist in der NBE selbst enthalten, die an die Stakeholder des Unternehmens als Informationsempfänger adressiert ist. Für die Bestimmung einer glaubwürdigen Botschaft bedient sich die Accountingforschung der Theorie des kommunikativen Handelns, welche vom deutschen Soziologen und Philosophen Jürgen Habermas entwickelt wurde (z.B. Lock, 2016, S. 422; Mazzotta et al., 2020, S. 1902). Habermas differenziert zwischen verschiedenen Geltungsansprüchen, die einer idealen Kommunikation auffindbar sein sollten: (a) Wahrheit, (b) Wahrhaftigkeit, (c) Angemessenheit und Verständlichkeit (Habermas, 2002, S. 63-64).

Bei der Wahrheit geht es darum, dass der Kommunikator in seiner Botschaft bei der objektiven Wahrheit bleibt. Die Wahrhaftigkeit kann als subjektive Wahrheit betrachtet werden und beschreibt die Eigenschaft, in seiner Botschaft aufrichtig zu sein.[2] Angemessenheit und Verständlichkeit bedeutet, dass das Unternehmen angemessen kommuniziert sowie eine möglichst einfache Sprache zu verwenden versucht (Habermas, 2002, S. 63-64; Reynolds & Yuthas, 2008, S. 53-54).

Wenngleich immer mehr Unternehmen NFI publizieren, ist die Glaubwürdigkeitslücke paradoxerweise gestiegen (Waddock & Goggins, 2011, S. 25). Demnach stellt das Erreichen von gesellschaftlicher Akzeptanz eine Herausforderung dar. Mangelnde Legitimation kann schlimmstenfalls dazu führen, dass die Umwelt bzw. die Stakeholder ein Unternehmen boykottieren, in der Öffentlichkeit negative Kritik entsteht oder gar rechtliche Konsequenzen (z.B. in Form von einer strengeren künftigen Gesetzgebung) für das Unternehmen zu befürchten sind (Spelthahn et al., 2009, S. 62).

2.5. Indikatoren für eine glaubwürdige Nachhaltigkeitsberichterstattung

Da die Glaubwürdigkeit ein komplexes Konstrukt darstellt, ist es notwendig im Rahmen dieses Kapitels auf mögliche Indikatoren für eine glaubwürdige NBE einzugehen. Die Erläuterungen orientieren sich an dem Glaubwürdigkeitsindex nach Mazzotta et al. (2020, S. 1907), welcher sich an den vier Geltungsansprüchen Wahrheit, Wahrhaftigkeit, Angemessenheit und Verständlichkeit von Habermas orientiert. In Abbildung 6 sind die Indikatoren für eine glaubwürdige NBE illustriert. Bei den genannten Aspekten handelt sich um Merkmale einer NBE, die bereits in der Forschung theoretisch bzw. empirisch untersucht wurden.

Der Rest des Kapitels widmet sich der Erklärung der einzelnen Aspekte aus Abbildung 6. Im empirischen Teil der Masterarbeit werden die Aspekte in messbare Variablen umgewandelt. Dieses Kapitel repräsentiert das argumentative Fundament für die ordinale Bewertung der im Glaubwürdigkeitsindex verwendeten Variablen.

2.5.1. Externe Prüfung (Wahrheit)

Um dem Misstrauen der Stakeholder entgegenzuwirken, beauftragen Unternehmen externe Prüfer, die die veröffentlichten Informationen der NBE kritisch würdigen (Spelthahn

[2] Ein Beispiel, was den Glaubwürdigkeitsverlust durch eine unwahrhaftige Aussage wiedergibt, ist der Abgasskandal des Volkswagen-Konzerns. Noch 2014 betonte Volkswagen in seiner NBE stets, dass eine umweltfreundliche Mobilität zentraler Bestandteil der Unternehmensstrategie sei (Blackwelder, Coleman, Colunga-Santoyo, Harrison & Wozniak, 2016, S. 1). Ein Jahr später wurde der Abgasskandal publik. Um die Grenzwerte verschiedener Autoabgase einzuhalten, bediente sich der Konzern einer Abschalteinrichtung, die die Abgaswerte auf dem Prüfstand verfälschten (Siano, Vollero, Conte & Amabile, 2017, S. 29-30). Die Aufrichtigkeit der Botschaft in der NBE, nämlich eine nachhaltige Mobilität voranzutreiben, kann im Nachhinein angezweifelt werden und führt zu einer Schädigung des Vertrauens der Stakeholder in künftige nachhaltigkeitsbezogene Zielsetzungen des Unternehmens.

Abbildung 6: Indikatoren für eine glaubwürdige NBE (eigene Darstellung)

et al., 2009, S. 66). Diverse wissenschaftliche Publikationen belegen, dass eine externe Prüfung die Glaubwürdigkeit in der NBE erhöhen (Hodge, Subramaniam & Stewart, 2009, S. 178; Romero, Fernandez-Feijoo & Ruiz, 2014, S. 491-493; Wong & Millington, 2014, S. 880-882).

In der Literatur wird bei den Anbietern der Prüfungsdienstleistung zwischen Accounting Sustainability Assurance Provider (ASAP) und Non-Accounting Sustainability Assurance Provider (NASAP) differenziert (Edgley, Jones & Atkins, 2015, S. 3). Die ASAPs werden hauptsächlich durch die Big-Four-Prüfungsgesellschaften, bestehend aus Ernst & Young (EY), PricewaterhouseCoopers (PwC), KPMG und Deloitte, repräsentiert. Diese haben sich bereits frühzeitig am Markt positioniert, um ihr Dienstleistungsportfolio in der Prüfung von Nachhaltigkeitsberichten zu erweitern (Ackers, 2009, S. 11; Manetti & Becatti, 2009, S. 293). Dem gegenüber stehen die NASAP, die vorrangig aus Ingenieurs- und Zertifizierungsunternehmen, auf Nachhaltigkeit spezialisierte Unternehmensberatungen oder auch weitere Praktiker (z.B. NGOs, Akademische Institutionen, etc.) bestehen (Perego & Kolk, 2012, S. 15). Naheliegend für ein Unternehmen ist es, einen ASAP für die externe Prüfung zu beauftragen, da sich dieser bereits im Rahmen der Jahresabschlussprüfung mit der operativen Tätigkeit des Unternehmens beschäftigt hat (Gillet, 2012, S. 81). Dennoch gibt es einige Vorzüge für die Beauftragung von NASAPs. Es wird bezweifelt, dass ASAP als Experten für das Handels- und Steuerrecht dasselbe Kompetenzprofil wie Ethiker, Soziologen oder Naturwissenschaftler (häufig vorkommende Qualifikationen von NASAPs) besitzen (Farooq & De Villiers, 2017, S. 86; Gray, 2000, S. 262-263). Gleichzeitig waren die ASAPs historisch gesehen häufiger in Skandale verwickelt, welches die Objektivität und Unabhängigkeit in Frage stellt (Farooq & De Villiers, 2017, S. 87). Ein bekanntes Beispiel ist der Fall Arthur Andersen, in dem die ehemalige Big-Five-Prüfungsgesellschaft wichtige Unterlagen aus der Konzernabschlussprüfung von Enron trotz angekündigter Ermittlungen vernichteten und damit die Ermittlungen der Justiz behinderten (Li, 2010, S. 37-41). Aufgrund des Kompetenzarguments sowie einer höher wahrgenommenen Objektivität bzw. Unabhängigkeit bevorzugen die externen Stakeholder die NASAPs (Wong & Millington, 2014, S. 880). Folglich wird im Kontext einer glaubwürdigen NBE ein NASAP im direkten Vergleich einem ASAP bevorzugt.

Der Umfang der externen Prüfung kann variieren. Dieser kann entweder spezifische NFI (z.B. bestimmte nur bestimmte Kennzahlen) oder den kompletten Nachhaltigkeitsbericht umfassen. Der Inhalt der Prüfung wird ex ante zwischen dem Unternehmen und der Prüfungsgesellschaft festgelegt. Um das Ziel einer glaubwürdigen NBE zu erreichen, sollte die externe Prüfung vollumfänglich und nicht selektiv erfolgen (Mazzotta et al., 2020, S. 1905).

Die Prüfungssicherheit kann als Wahrscheinlichkeitsbegriff aus dem Prüfungswesen verstanden werden, der maßgeblich zur Qualität des Prüfungsurteils beiträgt (Leffson, Lippmann & Baetge, 1969, S. 16). Nach Hagest (1975, S. 35-40) macht die Prüfungssicherheit darüber eine Aussage, wie hoch die Wahrscheinlichkeit ist, dass das getroffene Prüfungsurteil korrekt ist. Dem gegenüber steht das Prüfungsrisiko, wonach ein Prüfer ein bestätigendes Prüfurteil fällt, obgleich Informationen wesentlich falsch dargestellt werden (Marten, Quick & Ruhnke, 2007, S. 214). Demnach besteht eine inverse Beziehung zwischen Prüfungssicherheit und Prüfungsrisiko dar, die mathematisch wie folgt ausgedrückt werden kann: Prüfungssicherheit [in %] = 1 – Prüfungsrisiko [in %]. Wenn das Prüfungsrisiko gemindert wird, dann erfolgt eine Erhöhung der Prüfungssicherheit und vice versa (Stibi, 1995, S. 47). Diverse Prüfungshandlungen, wie z.B. Aufbau- und Funktionsprüfungen, analytische und einzelfallbezogene Prüfungshandlungen, helfen dabei, das Prüfungsrisiko zu mindern (Almeling, Flick & Scharr, 2020, S. 57-71). Das Ausmaß der Prüfungshandlungen erfolgt in Abhängigkeit von der verlangten Prüfungssicherheit (Leffson et al., 1969, S. 16). Es besteht ein positiver Zusammenhang zwischen der Prüfungssicherheit und der Korrektheit des Prüfungsurteils. Folglich steigt bei einer höheren Prüfungssicherheit die Verlässlichkeit bzw. Glaubwürdigkeit des Prüfungsurteils (Lubitzsch, 2008, S. 100). In der Praxis wird meist entweder der International Standards on Assurance Engagement 3000 (ISAE 3000) oder der AccountAbility 1000

(AA1000) als Prüfungsstandard angewendet (Walterbusch, Handzlik & Teuteberg, 2013, S. 311). Bei der Überprüfung der NBE definieren die Prüfungsstandards, welche Grundsätze ordnungsgemäßer Prüfung und prüferischer Durchsicht während des gesamten Prüfungsprozesses (Auftragsannahme, Auftragsdurchführung, Berichterstattung und Erstellung des Prüfungsurteils bzw. Bestätigungsvermerks) zu befolgen sind (Walterbusch et al., 2013, S. 314-315). Im Gegensatz zum ISAE 3000, der weitgehend mit dem Prüfungsstandard für die Jahresabschlussprüfung übereinstimmt, ist der AA1000 speziell für die NBE entwickelt worden (J. Smith, Haniffa & Fairbrass, 2011, S. 426-427). Im AA1000 nimmt die Wahrnehmung der Interessen der Stakeholder einen besonders hohen Stellenwert ein (Adams & Evans, 2004, S. 99) Der ISAE 3000 wird primär von ASAP verwendet und der AA1000 von NASAP (Farooq & De Villiers, 2017, S. 87). Im Prüfungsstandard ISAE 3000 wird zwischen Limited Assurance (mittlere Prüfungssicherheit) und Reasonable Assurance (hohe Prüfungssicherheit) unterschieden. Der Prüfungsstandard AA1000 differenziert zwischen High Level of Assurance (hohe Prüfungssicherheit) und Moderate Level of Assurance (mittlere Prüfungssicherheit) (Walterbusch et al., 2013, S. 314-315). Da eine hohe Prüfungssicherheit dem Wahrheitsanspruch näher kommt und wesentliche Falschdarstellungen in der NBE unwahrscheinlicher werden, ist im Zusammenhang mit einer glaubwürdigen NBE eine hohe Prüfungssicherheit einer mittleren Prüfungssicherheit vorzuziehen.

2.5.2. Eigenschaften der NBE (Wahrheit)

Es existieren zahlreiche Rahmenwerke, die den Unternehmen dabei helfen, Nachhaltigkeitsberichte zu erstellen (Siew, 2015, S. 181-182). Großer Popularität erfreut sich die Global Reporting Initiative (GRI). Die GRI, welche als NGO im Jahre 2002 in Amsterdam gegründet wurde, etablierte ein Rahmenwerk für die NBE. Dieses Rahmenwerk enthält zahlreiche Standards, die dem Unternehmen aufzeigen, über was und in welcher Weise die Berichterstattung erfolgen sollte. Dies hilft dem Unternehmen dabei, über soziale, ökonomische sowie ökologische Themen in strukturierter Art und Weise zu berichten (Aifuwa, 2020, S. 15). Durch die Orientierung an Standards wird es für Stakeholder einfacher, die NBE zu lesen und Inhalte mit anderen Unternehmen zu vergleichen (Michalczuk & Konarzewska, 2020, S. 84-85). Bei der Verwendung des GRI Rahmenwerks gibt es die Option „Kern“ und die Option „Umfassend“. Die Option „Umfassend“ entspricht einem höheren Anwendungsniveau, welches mehr Offenlegungsvorgaben (z.B. mehr nichtfinanzielle Leistungsindikatoren) enthält, als die Option „Kern“ (Mazzotta et al., 2020, S. 1905).[3] In einer Untersuchung von Lock und Seele (2016, S. 192) konnte empirisch nachgewiesen werden, dass die Verwendung des GRI Rahmenwerks zu einer glaubwürdigeren Berichterstattung führt. Zudem führte das Anwendungsniveau Option „Umfassend“ zu einer glaubwürdigeren NBE als die Option „Kern“.

Die Quantität der veröffentlichten NFI variiert von Unternehmen zu Unternehmen und hängt von der jeweiligen Branche ab (Boiral, 2013, S. 1048; Roca & Searcy, 2012, S. 109). Ein umfangreicher Nachhaltigkeitsbericht gibt nicht unbedingt Aufschluss darüber, ob ein Unternehmen ein breites thematisches Spektrum zum Thema Nachhaltigkeit abdeckt (Boiral, 2013, S. 1051-1061), oder dass es von dem Rezipienten als nachhaltig wahrgenommen wird (Helfaya, Whittington & Alawattage, 2018, S. 176). Nichtsdestotrotz korreliert die Quantität der NFI stark signifikant mit einem für die NBE entwickelten Qualitätsindex, der die Glaubwürdigkeit beinhaltet (Helfaya & Whittington, 2019, S. 530, 534). Lock und Seele (2016, S. 193) konnten einen direkten Zusammenhang zwischen der Länge des Nachhaltigkeitsberichts (gemessen in der Wortanzahl) und der Glaubwürdigkeit empirisch nachweisen. Gleichzeitig besagt die Theorie, dass die Quantität mit in die Informationsqualität einfließt (Beretta & Bozzolan, 2008, S. 342). Aus den empirischen und theoretischen Ausführungen ist demnach abzuleiten, dass eine hohe Quantität von NFI als Indikator für eine glaubwürdige NBE spricht.

Für die Verortung der NFI gibt es drei verschiedene Möglichkeiten, die in der Praxis angewendet werden (Michelon, Pilonato & Ricceri, 2015, S. 63): (a) die Veröffentlichung der sozialen und ökologischen Aspekte im Lagebericht des Konzern- bzw. Jahresabschlusses („Annual Report“), (b) die Veröffentlichung eines Geschäftsberichts, der ökonomische, soziale und ökologische Aspekte umfasst („Integrated Reporting“) oder (c) die Veröffentlichung eines separaten Nachhaltigkeitsberichts. Variante a) erlaubt im Gegensatz zur Variante b) ausschließlich die Veröffentlichung der NFI im Lagebericht, wohingegen bei Variante b) die NFI im gesamten Bericht verteilt sind (Mazzotta et al., 2020, S. 1906). Das Bedürfnis der Stakeholder nach einem Integrated Reporting, bei dem finanzielle und nicht-finanzielle Aspekte in einem Bericht zusammenhängend dargestellt werden, wächst stetig (Eccles & Krzus, 2010, S. 146-155). Beispielsweise kann sich ein Unternehmen fragen, ob das Programm zur Steigerung der Energieeffizienz zur Einsparung von CO2-Emissionen zu einer Kostenreduktion in der Produktion führt (Eccles & Saltzman, 2011, S. 59). Trotz der naheliegenden Vorteile konnte bislang keine empirische Evidenz die These untermauern, dass das Integrated Reporting zu einer glaubwürdigeren NBE führt (Lock & Seele, 2016, S. 193). Um trotzdem die Verortung der NFI mit der Glaubwürdigkeit der NBE in Verbindung zu bringen, kann die Quantität (s.o.) verwendet werden. Eine Veröffentlichung im Lagebericht führt i.d.R. zu einer geringen Quantität, das Integrated Reporting zu einer mittleren Quantität und ein separater Nachhaltigkeitsbericht zu einer hohen Quantität von NFI (Mazzotta et al., 2020, S. 1906). Ferner konnten Michelon et al. (2015, S. 72) zeigen, dass ein separater Nachhaltigkeitsbericht mit der Quantität positiv korreliert ist. Dies stützt die These von Mazzotta et al. (2020), dass die Verortung von NFI die Quantität und folglich auch die Glaubwürdigkeit beeinflusst.

[3] Neben dem GRI Rahmenwerk existieren noch weitere Rahmenwerke, wie z.B. der Deutsche Nachhaltigkeitskodex (DNK) und der Global Compact Index (Ayuso, Roca, Arevalo & Aravind, 2016, S. 553; Zwick, 2018, S. 63).

2.5.3. Genauigkeit (Wahrheit)

Das Vorhandensein eines Kapitels zur angewandten Methodik in der NBE gewährleistet, dass die Rahmenbedingungen der NBE adäquat kommuniziert werden. In einem Methodik-Kapitel sollten folgende Inhalte mindestens enthalten sein: (a) Angaben zu den enthaltenen Entitäten, (b) Neudarstellung von Informationen oder Änderungen in der Berichterstattung, (c) der Berichtszeitraum und (d) das allgemeine Vorgehen zur Bestimmung des Berichtsinhalts (Mazzotta et al., 2020, S. 1906)[4]. Mit der Offenlegung dieser allgemeinen Angaben ist sichergestellt, auf welche Gesellschaften und welchen Zeitraum sich die Informationen beziehen. Ferner wird allgemein der Ansatz zur Bestimmung des Berichtsinhalts formuliert, sodass für Stakeholder klar nachvollziehbar ist, wie bei der Berichterstellung vorgegangen wurde. Um möglichst glaubwürdig zu berichten, ist es erforderlich, dass die Stakeholder über die oben genannten Inhalte in Kenntnis gesetzt werden, damit die NFI entsprechend evaluiert werden können. Folglich ist eine glaubwürdige NBE durch das Vorhandensein eines Methodik-Kapitels gekennzeichnet.

Nachdem mit einem Methodik-Kapitel die Rahmenbedingungen für die NBE kommuniziert sind, geht es bei der Genauigkeit explizit um die Tiefe der offengelegten NFI (GRI, 2016, Std. 102, S. 13). Die GRI beschreibt die Genauigkeit wie folgt: "The reported information shall be sufficiently accurate and detailed for stakeholders to access the reporting organization's performance." (GRI, 2016, Std. 102, S. 13) Für eine gelungene NBE ist es somit notwendig, möglichst genaue Informationen zu liefern (Clarke, 2007, S. 241). Die Beschreibung von NFI kann auf qualitativer, quantitativer oder monetärer Weise erfolgen. Die Genauigkeit ist umso stärker, desto mehr der drei Möglichkeiten zur Beschreibung verwendet werden (Michelon et al., 2015, S. 67; Wiseman, 1982, S. 55). Helfaya et al. (2018, S. 174) fanden in einer Umfrage heraus, dass 86 % der Stakeholder die Genauigkeit als wichtiges Element der NBE ansehen. Dies unterstreicht die Bewertung von Mazzotta et al. (2020, S. 1903, 1906), die Genauigkeit als Indikator für eine glaubwürdige NBE einzustufen.

2.5.4. Wesentlichkeit (Wahrhaftigkeit)

Das Prinzip der Wesentlichkeit besagt, dass ein Unternehmen in der NBE zum einen nur über Themen berichten soll, welche erhebliche Auswirkungen auf ökologische, ökonomische oder soziale Aspekte aus der Unternehmensperspektive haben (z.B. Risiken für die Reputation des Unternehmens). Zum anderen müssen diese Themen die Beurteilungen und Entscheidungen der Stakeholder in besonderem Maße beeinflussen (GRI, 2016, Std. 101, S. 10). In Abbildung 7 skizziert diesen zweidimensionale Zusammenhang in einer Wesentlichkeitsmatrix. Die blauen Ovale stehen für verschiedene Themen, über die ein Unternehmen berichten kann. Ein Unternehmen sollte in jedem Fall über die drei Themen berichten, die sich im ersten Quadranten befinden. Diese Vorgehensweise ist notwendig, um eine Priorisierung der Themen vorzunehmen (GRI, 2016, Std. 101, S. 11).

Zur Themenbestimmung hat sich das Aufstellen einer Wesentlichkeitsmatrix weitgehend etabliert (Jones, Comfort & Hillier, 2016, S. 84). Es zeigt, dass ein Unternehmen sich angemessen mit der Themenschwerpunktsetzung beschäftigt. Da sich die wesentlichen Themen andauernd verändern können, ist eine regelmäßige Überprüfung auf deren Aktualität essentiell (Ankele & Grothe, 2019, S. 565). Mazzotta et al. (2020, S. 1906) bewerten das Aufstellen einer Wesentlichkeitsmatrix, die Aktualität der Wesentlichkeitsanalyse sowie die Ratifizierung der Wesentlichkeitsanalyse durch den Vorstand als glaubwürdigkeitssteigernde Einflussgrößen auf die NBE.

2.5.5. Stakeholderbeziehung (Wahrhaftigkeit)

Der Grad der Interaktion und der Teilhabe mit den Stakeholdern kann unternehmensabhängig verschieden ausfallen. Manetti (2011, S. 110-111) hat das Stufenleitermodell nach Arnstein (1969) in den Kontext der NBE gesetzt. Die erste Stufe des Stakeholder-Managements ist, dass sich das berichtende Unternehmen darüber im Klaren ist, wer genau die internen und externen Stakeholder sind. Es handelt sich also um eine reine Identifikation der Stakeholdergruppen. Sobald ein Unternehmen einen zweiseitigen Austausch bzw. einen Dialog mit den Stakeholdern zulässt (z.B. in Form einer öffentlichen Diskussion), ist die nächste Stufe der Partizipation erreicht (Manetti, 2011, S. 110). Die letzte Stufe der Partizipation ist das Stakeholder-Engagement. In dieser Stufe werden die Interessen der Stakeholder in Planungs- und Entscheidungsprozesse aktiv miteinbezogen (z.B. durch Interessensvertreter im Aufsichtsrat) (Manetti, 2011, S. 111). Eine hohe Einbeziehung der Stakeholder spricht dafür, dass sich ein Unternehmen ernsthaft über die Interessen der Anspruchsgruppen Gedanken macht und den gemeinsamen Austausch pflegt. Im besten Falle besitzen die Stakeholder sogar einen Einfluss in der Entscheidungsfindung. Es gilt folglich: Je höher der Partizipationsgrad der Stakeholder ist, umso glaubwürdiger ist die NBE (Mazzotta et al., 2020, S. 1907).

2.5.6. Organisationale Verankerung der Nachhaltigkeit (Wahrhaftigkeit)

In Unternehmen erfolgt i.d.R. die organisationale Verankerung der Nachhaltigkeit in einem Sustainability Komitee (Biswas, Mansi & Pandey, 2018, S. 522). Ein solcher (meist) multipersonaler Zusammenschluss kann in bestehende organisationale Strukturen integriert werden (z.B. in die Compliance-Abteilung). Speziell für die Nachhaltigkeit geschaffene organisationale Strukturen, wie z.B. die Organisation in einer eigenständigen Nachhaltigkeitsabteilung oder dezentrale Strukturen über verschiedene Abteilungen hinweg, sind auch möglich (Cucari, Esposito De Falco & Orlando, 2018, S. 255). Beide Formen besitzen den gleichen Zweck, nämlich definierte Nachhaltigkeitsziele in tatsächliche Maßnahmen umzusetzen. Diese Maßnahmen werden vom Su-

[4] Die Mindestinhalte wurden indirekt durch das Nachvollziehen der Inhaltsanalyse von Mazzotta et al. (2020, S. 1909) ermittelt.

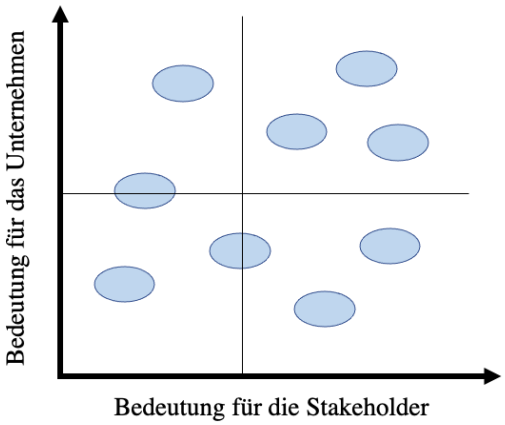


Abbildung 7: Wesentlichkeitsmatrix (in Anlehnung an Ankele & Grothe, 2019, S. 565)

stainability Komitee geplant, implementiert und überwacht (Liao, Luo & Tang, 2015, S. 414; Peters & Romi, 2015, S. 163). Die Rolle eines Sustainability Komitees in Bezug auf die Qualität von NFI ist in der Literatur umstritten. Sowohl Michelon und Parbonetti (2012) als auch Rupley, Brown und Marshall (2012) konnten keinen positiven Zusammenhang zwischen dem Vorhandensein eines Sustainability Komitees und der Qualität der NFI feststellen. Auf der anderen Seite gibt es empirische Ergebnisse, die auf eine positive Wirkung bezüglich der Qualität von NFI hindeuten (z.B. Amran, Lee & Devi, 2014; Arena, Bozzolan & Michelon, 2015; Helfaya & Moussa, 2017; Liao et al., 2015). Zum Beispiel wurden in der Studie von Liao et al. (2015, S. 409) 329 britische Unternehmen observiert. Gemachte Angaben zu den Emissionswerten von Treibhausgasen waren bei Vorhandensein eines Sustainability Komitees hinsichtlich der Qualität der NFI besser. Ungeachtet der objektiven Qualität von NFI ist die Glaubwürdigkeit insbesondere auch von der Wahrnehmung der Stakeholder abhängig (Helfaya et al., 2018, S. 184). Das Vorhandensein eines Sustainability Komitees steigert die Legitimität einer Unternehmung in den Augen der Stakeholder (Neu, Warsame & Pedwell, 1998, S. 265; Yunus, Elijido-Ten & Abhayawansa, 2016, S. 156). Da die Glaubwürdigkeit als Voraussetzung mit der Legitimität zusammenhängt (s. Kap. 2.4), kann das Vorhandensein eines Sustainability Komitees als glaubwürdigkeitssteigender Indikator eingeordnet werden (Lock & Schulz-Knappe, 2019, S. 2; Seele & Lock, 2015, S. 404). Mazzotta et al. (2020, S. 1906) bewertet eine speziell für die Nachhaltigkeit geschaffene organisationale Verankerung glaubwürdiger als eine integrierte Lösung.

2.5.7. Sustainable Development Goals (Wahrhaftigkeit)

Am 1. Januar 2016 wurden die Sustainable Development Goals (SDGs) von den Vereinten Nationen mit einer Laufzeit von 15 Jahren (bis 2030) auf den Weg gebracht. Im Unterschied zu den Millennium Development Goals, die von 2001 bis 2015 von Bedeutung waren, sind die SDGs für alle Länder dieser Welt und nicht nur für Entwicklungsländer konzipiert (Mio, Panfilo & Blundo, 2020, S. 3221). Sie können als globaler Konsens angesehen werden, welche Ziele mit einer nachhaltigen Entwicklung einhergehen. Damit werden die drei Säulen Wirtschaft, Soziales und Umwelt (s. Kap. 2.2) berücksichtigt (Sachs, 2012, S. 2206). Um eine erfolgreiche NBE zu gestalten, müssen die Nachhaltigkeitsziele in die Vision bzw. Strategie des Unternehmens eingebettet sein (Dilling, 2010, S. 22). Obwohl keine empirische Studien den Zusammenhang zwischen den SDGs und einer glaubwürdigen NBE untersucht haben, wird die Bezugnahme zu den anerkannten SDGs gemäß Mazzotta et al. (2020, S. 1903, 1906) als positiver Indikator für die Glaubwürdigkeit gewertet.

2.5.8. Einfachheit in der Kommunikation (Angemessenheit & Verständlichkeit)

> "A plain English document uses words economically and at a level the audience can understand. Its sentence structure is tight. Its tone is welcoming and direct. Its design is visually appealing. A plain English document is easy to read and looks like it's meant to be read." (SEC, 1998, S. 5)

Dieses Zitat verdeutlicht, dass die Informationsaufnahme für den Leser bzw. Stakeholder besser gelingt, wenn die Präsentation in einfacher Sprache erfolgt. Bisherige Studien aus dem Financial Accounting belegen, dass die Lesbarkeit bzw. die Verwendung einer einfachen Sprache die Glaubwürdigkeit maßgeblich beeinflussen (Rennekamp, 2012, S. 1319; Tan, YING WANG & Zhou, 2014, S. 273). Das Quasi-Experiment von Hoozée, Maussen und Vangronsveld (2019, S. 570) untermauert diese Erkenntnis im Kontext der NBE, wonach die wahrgenommene Glaubwürdigkeit durch eine gestiegene Lesbarkeit (bei nichtprofessionellen Lesern) induziert wurde. Demnach ist eine möglichst einfache Sprache für eine glaubwürdige NBE von Nöten (Mazzotta et al., 2020, S. 1906).

Die Verwendung von visuellen Darstellungen, wie z.B. Bildern, Grafiken oder Tabellen, sind bei der Informationsübermittlung hilfreich und vereinfachen damit die Kommu-

nikation. Kelly (1993, S. 148) untersuchte in einem Experiment, ob die Verarbeitung von Informationen durch visuelle Darstellungen beeinflusst wird. Die Verwendung von Tabellen und Grafiken konnte dabei die Dauer der Informationsverarbeitung im Vergleich zur bloßen Textdarstellung signifikant reduzieren. Des Weiteren begrüßen Stakeholder das Vorhandensein von visuellen Darstellungen (Helfaya et al., 2018, S. 179). Die Benutzung von visuellen Darstellungsmöglichkeiten sind aus dem Grund der einfacheren Informationsaufnahme für den Rezipienten positiv im Hinblick auf die Glaubwürdigkeit in der NBE zu betrachten (Mazzotta et al., 2020, S. 1906).

3. Regulatorischer Hintergrund

Die seit der Jahrtausendwende zunehmende Wahrnehmung in der europäischen Bevölkerung, dass privatwirtschaftliche Unternehmen immer stärker entgegen des Gemeinwohls agieren, hat sich im Zuge verschiedener Krisen (z.B. Finanz- und Klimakrise) in der öffentlichen Debatte manifestiert (Schweren & Brink, 2016, S. 177). Aufgrund des mangelnden Vertrauens der Stakeholder in die unternehmerischen Aktivitäten, hat sich der europäische Gesetzgeber für eine verpflichtende Berichterstattung über NFI entschieden (EU, 2014a, S. 1; Schweren & Brink, 2016, S. 177-178). Das legislative Resultat ist die CSR-RL, welche im November 2014 verabschiedet wurde (Stawinoga, 2017, S. 213). Die CSR-RL musste in allen europäischen Mitgliedsstaaten in die nationale Gesetzgebung einfließen und gilt für alle Geschäftsjahre, die nach dem 31.12.2016 beginnen (EU, 2014a, S. 8). Im Rahmen der Masterarbeit wird eine deutsche Stichprobe analysiert. Aus diesem Grund ist das CSR-RUG von Bedeutung. Das CSR-RUG orientiert sich nahezu 1:1 an dem Minimalkonsens der CSR-RL (Kajüter, 2017a, S. 137).

Damit ein Unternehmen sich den Vorgaben des CSR-RUG (§§ 289a-289e, 315a-315d HGB) unterwerfen muss, müssen folgende drei Voraussetzungen nach § 289b Abs. 1 Nr. 1-3 HGB kumulativ erfüllt sein: (1) Es handelt sich um eine große Kapitalgesellschaft i.S.d. § 267 Abs. 3 S. 1 HGB. (2) Es muss sich um eine kapitalmarktorientierte Gesellschaft i.S.d. § 264d HGB handeln. (3) Im Jahresdurchschnitt muss die Gesellschaft mehr als 500 Mitarbeiter besitzen. Diese Voraussetzungen gelten analog für kapitalmarktorientierte, haftungsbeschränkte Personengesellschaften i.S.d. § 264a HGB sowie Genossenschaften i.S.d. § 336 Abs. 2 S. 1 Nr. 2 HGB. Kreditinstitute sowie Versicherungen fallen unabhängig von der Kapitalmarktorientierung unter das CSR-RUG, sofern sie gemäß § 267 Abs. 3 S. 1 HGB als groß anzusehen sind und mehr als durchschnittlich 500 Mitarbeiter beschäftigen (§ 340a Abs. 1a S. 1 HGB, § 341a Abs. 1a S. 1 HGB). Mutterunternehmen i.S.d. § 290 HGB müssen dann den Berichtspflichten des CSR-RUGs nachkommen, wenn eine Kapitalmarktorientierung vorliegt, keine Befreiung i.S.d. § 293 Abs. 1 S. 1-2 HGB möglich ist und die Mitarbeiterzahl größer als 500 ist (Stawinoga, 2017, S. 217). Beim Gesetzesentwurf wurde die Anzahl der Unternehmen, die unter das CSR-RUG in Deutschland fallen, auf 548 geschätzt (222 einzelberichtspflichtige Kapitalgesellschaften, 171 konzernberichtspflichtige Mutterunternehmen sowie 155 Konzern- und einzelberichtspflichtige Mutterunternehmen) (Bundesregierung, 2017, S. 38).

Berichtspflichtige Unternehmen nach dem CSR-RUG sind gemäß § 289c Abs. 1 HGB (für Konzerne: § 315c Abs. 1 HGV i.V.m. § 289c Abs. 1 HGB) verpflichtet, eine nichtfinanzielle Erklärung (NFE) zu erstellen. Inhalt der NFE sind eindeutig im § 289c Abs. 2 Nr. 1-5 HGB bestimmt. Zu folgenden Themen muss das berichtende Unternehmen mindestens Angaben machen: (1) Umweltbelange, (2) Arbeitnehmerbelange, (3) Sozialbelange, (4) Achtung der Menschenrechte sowie (5) Bekämpfung von Korruption und Bestechung. Der § 289c Abs. 3 HGB konkretisiert die darzulegenden Angaben in der sog. Wesentlichkeitsformel. Demnach müssen NFI zu den Themen enthalten sein, die „[...] für das Verständnis des Geschäftsverlaufs, des Geschäftsergebnisses, der Lage der Kapitalgesellschaft sowie der Auswirkungen ihrer Tätigkeit [...] erforderlich sind." (§ 289c Abs. 3 HGB) Dazu gehört, dass das Unternehmen eine klare Strategie für die obigen Themen verfolgt, die dazugehörigen Maßnahmen schildert, die dazugehörigen Governance-Strukturen darlegt sowie die internen Prüfprozesse erklärt (DNK, 2018, S. 17). Zudem sind wesentliche Risiken, die mit der Geschäftstätigkeit einhergehen, offenzulegen und nichtfinanzielle Leistungsindikatoren zu verwenden (§ 289c Abs. 3. Nr. 3-5 HGB). Demnach geht das CSR-RUG thematisch und auch von den geforderten Angaben weit über den zuvor geltenden § 289 Abs. 3 HGB, der sich nur auf nichtfinanzielle Leistungsindikatoren bezog, hinaus. Allerdings gilt im CSR-RUG der „Comply-or-Explain"-Ansatz gemäß § 289c Abs. 4 HGB. Demnach müssen die Angaben gemäß der obigen Bestimmungen erfolgen oder es muss plausibel dargelegt werden, weswegen keine Anstrengungen erfolgt sind (Szabó & Sørensen, 2015, S. 332). Ferner erlaubt der Gesetzgeber, sofern die Offenlegung von NFI mit erheblichen Nachteilen i.S.d. § 289e Abs. 1 HGB für das Unternehmen verbunden ist, eine Befreiung von Publizitätspflicht. Allerdings handelt es sich bei dieser Regelung nicht um eine Verzichtsmöglichkeit. Vielmehr wird die Offenlegung auf einen späteren Zeitpunkt verschoben (Heckler & Bröcker, 2017, S. 764).

Ein berichtspflichtiges Unternehmen hat drei Möglichkeiten zur Offenlegung der NFE: (1) Die Publikation im Lagebericht, (2) die Erstellung einer gesonderten NFE, die zeitgleich mit dem Lagebericht nach § 325 HGB im Bundesanzeiger veröffentlich wird oder (3) die gesonderte NFE wird im Lagebericht referenziert und es wird auf die Internetseite verwiesen (DNK, 2018, S. 9). Das Vorhandensein einer NFE wird lediglich vom Wirtschaftsprüfer festgestellt (DNK, 2018, S. 9). Eine externe Prüfung ist demnach im CSR-RUG nicht vorgesehen. Im Falle einer AG muss der Aufsichtsrat die NFE, welche vom Vorstand erstellt wird, nach § 171 Abs. 1 S. 4 AktG inhaltlich prüfen. Folglich wird nur eine interne inhaltliche Prüfung vorgenommen. Das CSR-RUG empfiehlt im § 289d HGB explizit die Verwendung eines Rahmenwerks. Sofern kein Rahmenwerk verwendet wird, ist dies zu erläutern. Ferner wurden im Zuge der Gesetzesreform die Sanktionsmöglichkeiten, z.B. durch höhere Bußgelder, angepasst

(§ 331 ff. HGB).

Die verbindliche Verpflichtung zur Offenlegung einer NFE kann als Meilenstein in der NBE angesehen werden (Rehbinder, 2014, S. 25). Insbesondere ist hervorzuheben, dass Unternehmen, die sich jahrelang nicht in der Pflicht zur Offenlegung von NFI sahen, nun durch die Gesetzesreform dazu gezwungen werden (Rehbinder, 2014, S. 27). Aus der rechtswissenschaftlichen Perspektive gibt es jedoch einige Kritikpunkte an der CSR-RL. Aufgrund der Wesentlichkeitsformel ist die NFE nicht als umfassender Nachhaltigkeitsbericht anzusehen (Kajüter, 2017b, S. 620-621). Beispielsweise bleibt die Berichtserstattung über nachhaltigkeitsbezogene Themen entlang der Lieferketten weiterhin freiwillig (Grob, Sydow & Heinz, 2020, S. 33-34). Zudem wird in Fachkreisen negativ angemerkt, dass der Adressatenkreis berichtspflichtiger Unternehmen zu klein ausfällt und damit die Wirkungskraft der CSR-RL begrenzt ist (Stawinoga, 2017, S. 217). Trotzdem wird erwartet, dass es zu Ausstrahlungseffekten auf Unternehmen kommt, die mit den berichtspflichtigen Unternehmen wirtschaftlich verbunden sind. Somit werden mittelbar auch kleine und mittelständische Unternehmen betroffen sein (Böcking & Althoff, 2017, S. 249). Außerdem wird die fehlende Verpflichtung zur Verwendung eines Rahmenwerks angeprangert. Das Ziel, eine Vergleichbarkeit und Transparenz zwischen den EU-Mitgliedstaaten in der NBE zu erreichen, wird somit nicht ernsthaft berücksichtigt in der CSR-RL (EU, 2014a, S. 1-2).

4. Literaturüberblick

In der Literatur ist der Effekt einer verpflichtenden NBE umstritten. Einige Wissenschaftler sind der Ansicht, dass eine obligatorische NBE nicht angemessen angesichts der hohen Heterogenität der Unternehmen ist. Demnach wird eine Einheitslösung als nicht zielführend angesehen (ICC, 2015, S. 2). Zudem kann der Innovationsgrad der NBE durch ein enges gesetzliches Korsett ausgebremst werden (Hahn & Kühnen, 2013, S. 7; ICC, 2015, S. 2). Auf der anderen Seite argumentieren die Fürsprecher der Berichtspflicht, dass durch eine gesetzliche Verankerung der Wunsch der Stakeholder nach mehr NFI nachgekommen wird und somit eine bessere Kontrolle der Unternehmen möglich ist (Hess, 2008, S. 451). Zudem werden die Berichtsverweigerer gezwungen, NFI zu publizieren (Kolk & Pinkse, 2010, S. 25).

Neben dieser theoretischen Debatte existieren zahlreiche wissenschaftliche Artikel, die den Effekt einer verpflichtenden Berichterstattung auf die NBE empirisch untersucht haben (Gulenko, 2018, S. 8). Während eindeutig belegt ist, dass nach einer Berichtspflicht die Anzahl der publizierenden Unternehmen steigt (z.B. Dong & Xu, 2016; Dumitru, Dyduch, Guşe & Krasodomska, 2017) und ebenso die Quantität der NFI zunimmt (z.B. Haji, 2013; Kerret, Menahem & Sagi, 2010), ist der Effekt auf die Qualität nicht eindeutig. Das Verständnis, was Qualität im Kontext der NBE eigentlich bedeutet, unterscheidet sich in der Literatur erheblich. Dies spiegelt sich in den unterschiedlichen Indices zur Messung der Qualität wider (Helfaya & Whittington, 2019, S. 530). Die Glaubwürdigkeit kann als eine Dimension der Qualität angesehen werden (z.B. Helfaya et al., 2018, S. 186; Mion & Loza Adaui, 2019, S. 11). Im Zusammenhang mit dem angestrebten Forschungsziel der Masterarbeit, nämlich die Auswirkungen der CSR-RL auf die Glaubwürdigkeit zu untersuchen, wird im Folgenden auf die Studien eingegangen, die sich explizit auf die Glaubwürdigkeit beziehen und den Unterschied zwischen einer freiwilligen und verpflichtenden NBE untersuchen.

Lock und Seele (2016, S. 189) untersuchen in ihrer empirischen Studie 237 Nachhaltigkeitsberichte aus den Jahren 2011-2014 aus insgesamt elf europäischen Ländern. Es wird die Hypothese aufgestellt, dass Unternehmen aus Ländern mit einer verpflichtenden NBE eine höhere Glaubwürdigkeit aufweisen als in Ländern ohne regulatorische Vorgaben. Diese Erwartung wird auf Basis von empirischen Ergebnissen hergeleitet. Für die Bestimmung der Glaubwürdigkeit wurde ein Index entwickelt, der aus insgesamt 93 Variablen besteht (Lock & Seele, 2016, S. 195). Die Studie orientiert sich an den vier Geltungsansprüchen – nämlich Wahrheit, Wahrhaftigkeit, Angemessenheit und Verständlichkeit – nach Habermas (Lock & Seele, 2016, S. 187). Die Daten für den verwendeten Glaubwürdigkeitsindex wurden mittels einer Inhaltsanalyse bestimmt (Lock & Seele, 2016, S. 186). Für die Überprüfung, ob eine verpflichtende NBE zu einer glaubwürdigeren NBE führt, wurden die Länder mit gesetzlicher Regulierung (Frankreich, Spanien) mit den Ländern ohne gesetzliche Regulierung verglichen. Mittels eines t-Tests konnte kein statistisch signifikanter Unterschied zwischen den beiden Gruppen festgestellt werden. Allerdings zeigte sich ein statistisch signifikanter Unterschied zwischen den Ländern Frankreich und Spanien, was verdeutlicht, dass die jeweilige Ausgestaltung der nationalen Gesetzgebung maßgeblich für eine potentielle Glaubwürdigkeitssteigerung ist (Lock & Seele, 2016, S. 192-193). Adaui (2020, S. 1) untersucht in seiner Studie u.a. die Glaubwürdigkeit der NBE von 27 peruanischen Unternehmen über die Jahre 2014-2016. Als Teilaspekt der Qualität fließt die Glaubwürdigkeit mit einer Gewichtung von 31 % in den Qualitätsindex mit ein (Adaui, 2020, S. 7). Bei den neun Variablen zur Messung der Glaubwürdigkeit handelt es sich ausschließlich um dichotome Variablen, wie z.B. „Standards“, „Assurance“ oder „Accuracy“ (Adaui, 2020, S. 6-7). Wenngleich die deskriptiven Ergebnisse für einen Anstieg der Glaubwürdigkeit im Jahr der verpflichtenden NBE (2016) sprechen, konnte dieser Zusammenhang nicht mittels Inferenzstatistik (Wilcoxon-Vorzeichen-Rang-Test) belegt werden (Adaui, 2020, S. 10, 13).

Auf der anderen Seite existiert jedoch auch empirische Evidenz für eine Verbesserung der Glaubwürdigkeit nach der Einführung einer verpflichtenden NBE. Hąbek und Wolniak (2016, S. 403) untersuchten in ihrer Studie sechs verschiedene EU-Länder (Dänemark, Schweden, Frankreich, Großbritannien, Niederlande, Polen). Mittels einer Inhaltsanalyse wurde die Glaubwürdigkeit (als Teilaspekt der Qualität) in den Kategorien „Readability“, „Basic reporting principles“, „Quality of data“, „Stakeholder dialogue outcomes“, „Feedback“ und „Independent verification“ von 507 Nachhaltig-

keitsberichten aus dem Jahr 2012 analysiert (Hąbek & Wolniak, 2016, S. 406). Je nach Erfüllungsgrad wurden die Variablen mit einer Punktzahl zwischen null und vier bewertet (Hąbek & Wolniak, 2016, S. 409). Mit Hilfe des Mann-Whitney-U-Tests konnte ein statistisch signifikanter Unterschied zwischen den Ländern mit verpflichtender NBE und ohne gesetzliche Verankerung festgestellt werden (Hąbek & Wolniak, 2016, S. 412-413). Im Gegensatz zu den bisher vorgestellten Studien, handelt es sich bei der empirischen Untersuchung von Mion und Loza Adaui (2019) um eine Zwei-Zeitpunkte-Betrachtung vor und nach der Einführung der CSR-RL. Damit handelt es sich um eine Studie, die dem Forschungsvorhaben der Masterarbeit am nächsten ist. Die Stichprobe der Studie umfasst insgesamt 36 italienische aus dem FTSE MIB und 30 deutsche Unternehmen aus dem DAX (Mion & Loza Adaui, 2019, S. 12). Die Glaubwürdigkeit stellt eine von drei Dimensionen der Qualität der NBE dar. Insgesamt werden sieben dichotome Variablen - nämlich u.a. „Adoption of guidelines“, „Independent verification“, „Stakeholder engagement“ - verwendet, um die Glaubwürdigkeit zu messen (Mion & Loza Adaui, 2019, S. 10-11). Mittels eines Wilcoxon-Vorzeichen-Rang-Tests konnte empirisch nachgewiesen werden, dass die Glaubwürdigkeit nach Einführung der CSR-RL sowohl für die deutschen als auch für die italienischen Unternehmen signifikant gestiegen ist. Dieses Resultat konnte auch nach Exklusion der acht First-time-Reporter aufrechterhalten werden und bestätigt die Robustheit des empirischen Befunds (Mion & Loza Adaui, 2019, S. 15-16).

Anhand der vorausgegangen Ausführungen wird deutlich, dass es gegensätzliche theoretische Argumente für und wider einer verpflichtenden NBE gibt. Dies spiegelt sich auch in den unterschiedlichen empirischen Ergebnissen wider. Zudem ist die nationale Ausgestaltung der Berichtspflicht entscheidend, ob es zu einer Glaubwürdigkeitssteigerung in der NBE kommt oder nicht. Der Literaturüberblick von Gulenko (2018, S. 11) bestätigt, dass insbesondere im Rahmen der CSR-RL und dessen Auswirkung auf die Glaubwürdigkeit weiterer Forschungsbedarf besteht. Ziel des Forschungsvorhabens ist es, weitere empirische Evidenz am Beispiel der CSR-RL für oder wider eine verpflichtende NBE zu schaffen.

5. Hypothesenherleitung

Die pragmatische Legitimität besagt, dass ein Unternehmen aus reinem Eigeninteresse nach Legitimation strebt. Gründe hierfür waren bspw., dass der Zugang zu Ressourcen (wie z.B. Kapital oder talentierten Mitarbeitern) erleichtert wird (s. Kap. 2.3). Die Nichteinhaltung der CSR-RL wird sanktioniert und führt zu einer Geldstrafe (s. Kap. 3). Dies wiederum führt zu einer Reduzierung des Gewinns und widerstrebt dem eigenen Interesse des Unternehmens. Folglich müssen die gesetzlichen Mindestanforderungen eingehalten werden, um künftige Zahlungsströme nicht durch Strafzahlungsaufwendungen zu mindern. Zudem wird durch die Einhaltung der Gesetze einer negativen Medienberichterstattung und somit einem möglichen Reputationsverlust vorgebeugt. Demnach ist es aus eigenen strategischen Erwägungen sinnvoll, den strengeren gesetzlichen Bestimmungen der CSR-RL Folge zu leisten. In der Vier-Stufen-Pyramide adressiert das Argument der pragmatischen Legitimität die ökonomische sowie die rechtliche Verantwortung eines Unternehmens.

Ein Unternehmen gilt nur dann als moralisch legitimiert, wenn die Erwartungen der Umwelt mit dem Tun des Unternehmens übereinstimmen (s. Kap. 2.3). Die moralische Legitimität zielt daher auf die ethische und philanthropische Verantwortung eines Unternehmens ab, da die Erwartungen der Umwelt über das Gesetz hinausgehen. Um eine hohe moralische Legitimität zu erreichen, ist es für ein Unternehmen folglich nicht ausreichend, sich auf die ökonomische und rechtliche Verantwortung zu beschränken. Um moralische legitimiert zu sein, muss ein Unternehmen eine glaubwürdige NBE anstreben, da der gemeinsame Austausch mit den Stakeholdern über die NBE als Kommunikationsmedium erfolgt (s. Kap. 2.2 und Kap. 2.4). Die Indikatoren für eine glaubwürdige NBE wurden im Kapitel 2.5 dargelegt. Die CSR-RL tangiert (teilweise) die Indikatoren für eine glaubwürdige NBE, in dem bspw. die Mindestinhalte der NBE erstmalig definiert wurden (s. Kap. 5). Unternehmen, die z.B. vor der CSR-RL nur über die Säule Umwelt berichtet haben, müssen nun auch über die Säule Soziales berichten.

Zusammengefasst bedeuteten die obigen Ausführungen, dass aufgrund der pragmatischen Legitimität davon auszugehen ist, dass sich die neuerdings berichtspflichtigen Unternehmen an die Vorgaben der CSR-RL halten werden. Außerdem tangiert die CSR-RL die Indikatoren für eine glaubwürdige NBE, was eine Glaubwürdigkeitssteigerung in der NBE erwarten lässt. Die gestiegene Glaubwürdigkeit in der NBE führt mittelbar zu einer höheren moralischen Legitimität. Ferner wurde im Kapitel 4 aufgezeigt, dass die empirischen Ergebnisse ambivalent sind. Nichtsdestotrotz spricht die Studie von Mion und Loza Adaui (2019), die dem Forschungsvorhaben ähnelt, für eine glaubwürdigere NBE. Aus der vorangegangenen theoretischen und empirischen Argumentation wird folgende Hypothese hergleitet: Die Einführung der CSR-RL führt zu einer glaubwürdigeren NBE.

6. Empirische Untersuchung: Inhaltsanalyse

Das Forschungsziel der Masterarbeit ist es, herauszuarbeiten, ob die CSR-RL zu einer glaubwürdigeren NBE führt. Um die aufgestellte Hypothese, dass die CSR-RL die Glaubwürdigkeit in der NBE steigert, zu untersuchen, wird im Folgenden die Herangehensweise erläutert: Zunächst wird im Kapitel 6.1 auf die Forschungsmethodik der Inhaltsanalyse eingegangen. Das Kapitel 6.2. erörtert, warum die MDAX-Unternehmen untersucht werden, weshalb manche Unternehmen nicht betrachtet werden und es erfolgt eine Beschreibung der gewählten Stichprobe. In Anschluss daran wird im Kapitel 6.3 ausführlich der verwendete Glaubwürdigkeitsindex von Mazzotta et al. (2020) vorgestellt. An dieser Stelle wird ausgiebig auf den Kodierungsrahmen eingegangen. Die

Messung der Glaubwürdigkeit anhand des genannten Glaubwürdigkeitsindex erfolgt sowohl für die Geschäftsjahre vor als auch nach der Einführung der CSR-RL. Demnach werden die Geschäftsjahre, die nach dem 31.12.2016 beginnen mit dem vorausgegangenen Geschäftsjahr verglichen. Die statistische Auswertung des ex-ante-ex-post-Vergleichs erfolgt mit Hilfe von deskriptiver und induktiver Statistik in den Kapiteln 6.5.1 sowie 6.5.2. In der Literatur sind Einflussgrößen bekannt, die die Glaubwürdigkeit beeinflussen können. Die relevanten Kontrollvariablen werden zunächst im Kapitel 6.4 erläutert und definiert. Diese fließen in die induktive Statistik des Kapitels 6.5.2 mit ein. Mit Hilfe einer linearen multiplen Regressionsanalyse wird für mögliche Störeinflüsse kontrolliert. Im Kapitel 6.5.3 werden anschließend die Ergebnisse diskutiert und im Kapitel 6.6 werden die Limitationen des Forschungsvorhabens dargestellt.

6.1. Forschungsmethodik

Im Rahmen der Masterarbeit wird zur Messung der Glaubwürdigkeit eine Inhaltsanalyse durchgeführt. Die Inhaltsanalyse ist definiert als „a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use." (Krippendorff (2004, S. 18)) Zweck dieser Forschungsmethodik ist also, aus qualitativen Informationen quantitative Daten abzuleiten. Wie in Kapitel 4 zu sehen ist, handelt es sich bei dieser Forschungsmethodik um eine übliche wissenschaftliche Verfahrensweise in der Accountingforschung, um Daten zu generieren. Zu unterscheiden ist die formale qualitative Inhaltsanalyse von der evaluativen qualitativen Inhaltsanalyse (Schreier (2014, S. 8-13)). Die formale qualitative Inhaltsanalyse beschränkt sich auf das Erfassen und Beschreiben von Informationen unter formalen und strukturellen Gesichtspunkten (z.B. Wortanzahl, Lesbarkeitsformeln, etc.) (Schreier (2014, S. 12)). Demgegenüber bewertet die evaluative qualitative Inhaltsanalyse die Qualität der Inhalte. Dahinter steht die Annahme, dass manche Informationen nützlicher sind als andere (Hooks & van Staden, 2011, S. 200). Beide Varianten der Inhaltsanalyse haben gemeinsam, dass ein Kodierungsrahmen mit relevanten Kategorien geschaffen wird und anschließend die zu bewertenden Daten anhand der vorher festgelegten Kodierung bewertet werden (Schreier, 2014, S. 20). Bei dem verwendeten Glaubwürdigkeitsindex von Mazzotta et al. (2020), der im Kapitel 6.3 erläutert wird, werden beide Varianten der Inhaltsanalyse verwendet. Es gibt Kategorien, die sich auf formale bzw. strukturelle Aspekte beziehen (z.B. Seitenlänge, Anzahl visueller Seiten, Lesbarkeitsformel nach Flesch). Anderseits bestehen im Glaubwürdigkeitsindex Kategorien, die ordinal im Sinne einer evaluativ qualitativen Inhaltsanalyse bewertet werden (z.B. Anbieter der Prüfungsleistung, Organisationale Verankerung der Nachhaltigkeit, etc.). Demnach folgt diese Vorgehensweise der Logik, dass das Vorliegen eines Merkmals nützlicher ist als ein anderes Merkmal. Da der Großteil der Kategorien der evaluativen qualitativen Inhaltsanalyse folgt, wird der Glaubwürdigkeitsindex der „Meaning-oriented Approach: Interpretative Content Analysis" zugeordnet (Helfaya et al., 2018, S. 167; Mazzotta et al., 2020, S. 1905). Im Gegensatz zu rein formal qualitativen Inhaltsanalysen ist demnach ein größerer Interpretationsspielraum in der Bewertung gegeben (M. Smith & Taffler, 2000, S. 638).

6.2. Stichprobe

Zunächst wird die Auswahl der Grundgesamtheit der Stichprobe erläutert, während anschließend eine Erklärung erfolgt, weswegen manche Unternehmen nicht in die Stichprobe mitaufgenommen werden konnten.

6.2.1. Begründung der gewählten Stichprobe

In der vorliegenden Masterarbeit wurden die Unternehmen des MDAX betrachtet. Während der DAX die 30 größten deutschen Unternehmen nach Marktkapitalisierung und Börsenumsatz abbildet, beinhaltet der MDAX die 50 nächstgrößeren Unternehmen. Der MDAX wurde 1996 von dem Unternehmen Deutsche Börse AG entwickelt und ist seit dem Teil der DAX-Familie, die aus dem DAX, MDAX sowie dem SDAX besteht (QONTIGO, 2021).

In der empirischen Accountingforschung hat es sich etabliert, die Stichprobenauswahl an einem Aktienindex anzulehnen (Aggarwal, 2013, S. 54-55). Dies birgt den Vorteil, dass die Unternehmen nach objektiven Kriterien in die Stichprobe einfließen und keiner Selektionsverzerrung im Auswahlprozess unterliegen.

Darüber hinaus beschränkt sich die Stichprobe bewusst auf eine EU-Nation Deutschland. Damit kann auf einen zwischenstaatlichen Vergleich der Umsetzung der CSR-RL in die nationale Gesetzgebung verzichtet werden. Ferner muss im Rahmen der empirischen Untersuchung nicht auf kulturelle Unterschiede zwischen den Ländern geachtet werden, was ebenso einen Vorzug einer „Ein-Land-Betrachtung" darstellt. Kulturelle Einflüsse besitzen nachweislich einen Einfluss auf die NBE (Gallego-Álvarez & Ortas, 2017, S. 351).

Daneben war vor dem CSR-RUG die NBE für deutsche Unternehmen weitestgehend freiwillig, sodass ein ex-ante-ex-post-Vergleich vor und nach der Gesetzesreform eine isolierte Betrachtung der Gesetzesreform erlaubt. Ein Zweijahresvergleich ist insofern gerechtfertigt, weil die Wirkungskraft durch die neue Gesetzgebung bereits unmittelbar nach dessen Inkrafttreten zu erwarten ist (Mion & Loza Adaui, 2019, S. 9).

Außerdem konnte vom Autor in der Literatur keine vergleichbare Stichprobe aufgefunden werden, die einen ex-ante-ex-post-Vergleich nach Einführung der CSR-RL untersucht hat. Neben den untersuchten DAX-Unternehmen in der empirischen Studie von Mion und Loza Adaui (2019), erfolgt demnach eine Erweiterung der Stichprobe, was die Robustheit der Ergebnisse nochmal verstärken kann.

6.2.2. Auswahlprozess

Es wurden jene MDAX-Unternehmen in die Analyse inkludiert, die vor der Indexanpassung im September 2018 im MDAX enthalten waren. Hierfür wurde die historische Entwicklung des MDAX nachvollzogen und jene Unternehmen

in die Stichprobe aufgenommen, die im August 2018 im Index waren (DAX-Indices, 2021, S. 8-14). Damit ergibt sich eine Grundgesamtheit von 50 Unternehmen. Ein Unternehmen unterliegt den Vorschriften des CSR-RUG gemäß § 289b HGB nur dann, wenn die einschlägigen Voraussetzungen kumulativ erfüllt sind. Laut § 267 Abs. 3 S. 2 HGB gilt, dass eine kapitalmarktorientierte Kapitalgesellschaft stets als große Kapitalmarktgesellschaft einzustufen ist. Da alle Unternehmen des MDAX kapitalmarktorientiert sind, sind die Voraussetzungen des § 289b Abs. 1 Nr. 1, 2 HGB erfüllt. Das Überschreiten der 500 Beschäftigten gemäß § 289b Abs. 1 Nr. 3 HGB trifft bis auf eine Ausnahme auf alle Unternehmen zu. Die Deutsche Euroshop AG ist eine Beteiligungsgesellschaft, die die Voraussetzung von mindestens 500 Beschäftigten nicht erfüllt. Damit wird die Deutsche Euroshop AG aus der Stichprobe exkludiert. Ferner gibt es vier Unternehmen, die aufgrund ihres Unternehmenssitzes im Ausland nicht dem deutschen CSR-RUG nach § 289b HGB unterliegen. Für drei Unternehmen ist das luxemburgische (Aroundtown S.A., Grand City Properties S.A., RTL Group S.A.) und für ein Unternehmen das niederländische (Airbus SE) Handelsrecht anzuwenden. Auf die Einbeziehung dieser Unternehmen wurde verzichtet, da ausschließlich der isolierte Effekt der deutschen Gesetzesreform betrachtet wird. Die NBE der Axel Springer SE, der WACKER Chemie AG sowie die Dürr AG beinhalten zeitliche Inkonsistenzen. Beispielsweise berichten die ersten beiden genannten Unternehmen nur alle zwei Jahre zum Thema Nachhaltigkeit und bei der Dürr AG erfolgt die NBE abweichend zum Geschäftsjahr. Zeitliche Inkonsistenzen erschweren den Vergleich zwischen den Unternehmen und es kann nicht mehr sinnvoll für Störeinflüsse kontrolliert werden. Folglich wurden die drei Unternehmen aus der Stichprobe entfernt. Die Bewertung der Glaubwürdigkeit für die Uniper SE kann aufgrund unzureichender Datenlage nicht vorgenommen werden. Online verfügbar ist lediglich ein Kurzbericht und in diesem wird explizit auf die vollständige Version für detaillierte Informationen hingewiesen. Nach Rücksprache mit der Senior Sustainability Managerin des Unternehmens stellte sich heraus, dass die detaillierten Informationen in der Vergangenheit im Internet abrufbar waren. Diese sind jedoch nicht mehr verfügbar.

6.2.3. Beschreibung der gewählten Stichprobe

Nach Abzug aller exkludierten Unternehmen beträgt die gesamte Stichprobengröße 41 (N = 41). Alle in die Inhaltsanalyse mit einbezogenen Unternehmen sind in Tabelle 2 dargestellt:

In Abbildung 8 ist die Branchenverteilung der betrachteten Unternehmen visuell in einem Kuchendiagramm dargestellt.

Zu erkennen ist, dass die beiden Branchen Maschinenbau, Bauwesen, Verkehr & Logistik sowie Finanzen über die Hälfte (51 %) der Gesamtstichprobe ausmachen. Gefolgt von den Branchen Handel & Konsum (17 %) sowie Chemie & Pharmazie (15 %) bilden die Branchen Medien (10 %) sowie Energie & Rohstoffe (7 %) das Schlusslicht.

6.3. Messung der Glaubwürdigkeit

Während im Kapitel 2.5 die Indikatoren für eine glaubwürdige NBE und deren ordinale Bewertung erläutert wurde, erfolgt in diesem Kapitel eine Beschreibung des Kodierungsrahmens. Alle diese Variablen sind notwendig, um das komplexe Konstrukt der Glaubwürdigkeit zu quantifizieren. Die Messung der einzelnen Variablen erfolgt in enger Anlehnung an Mazzotta et al. (2020). Sofern es bewusste Abweichungen in der Bewertung im Vergleich zu den Autoren gibt, wird darauf explizit hingewiesen. Für die Inhaltsanalyse wurden die Geschäftsberichte sowie - sofern vorhanden - der gesonderte Nachhaltigkeitsbericht zur Bestimmung der Glaubwürdigkeit verwendet.[5]

Die minimale Punktzahl des Glaubwürdigkeitsindex (GW) ist null und die maximal zu erreichende Punktzahl beträgt 21. Der Glaubwürdigkeitsindex setzt sich aufgeschlüsselt nach den Dimensionen von Habermas wie folgt zusammen: (1) Wahrheit (W) mit zehn Punkten (48 % der Gesamtpunktzahl), (2) Wahrhaftigkeit (WH) mit neun Punkten (43 % der Gesamtpunktzahl) sowie (3) Angemessenheit & Verständlichkeit (AV) mit zwei Punkten (9 % der Gesamtpunktzahl). Damit fließen die Dimensionen Wahrheit und Wahrhaftigkeit in etwa mit der gleichen Punktzahl in den Glaubwürdigkeitsindex ein, während die Dimension Angemessenheit & Verständlichkeit eine eher untergeordnete Gewichtung einnimmt.

Tabelle 3 illustriert die berechneten standardisierte Cronbach's Alpha-Werte. Mit Ausnahme der Dimensionen Wahrhaftigkeit und Angemessenheit & Verständlichkeit nach der CSR-RL sind die standardisierten Cronbach's Alpha-Werte größer als 0,7. Insbesondere zeigen die standardisierten Cronbach's Alpha-Werte für den Glaubwürdigkeitsindex, dass es sich um ein reliables Messinstrument handelt.

In Tabelle 4 sind alle Variablen mit den möglichen Ausprägungen und der jeweiligen Kodierung der Merkmalsausprägung aufgeführt.

6.3.1. Externe Prüfung (Wahrheit)

Beim *Anbieter* der Prüfungsleistung wird zwischen keiner externen Prüfung, einer externen Prüfung durch einen ASAP sowie einer externen Prüfung durch einen NASAP unterschieden. Sofern das betrachtete Unternehmen sich keiner externen Prüfung unterzog, wird kein Punkt vergeben. Bei der Begutachtung durch einen ASAP, d.h. durch eine Wirtschaftsprüfungsgesellschaft, wird ein Punkt vergeben, wohingegen die Prüfung durch einen NASAP mit zwei Punkten honoriert wird. Im wissenschaftlichen Artikel von Mazzotta et al. (2020, S. 1907) wird hingegen nur ein Punkt vergeben, wenn die Prüfung durch einen NASAP erfolgt. Bei der Analyse italienischer Unternehmen ergibt diese Herangehensweise in der Punktevergabe Sinn, da die nationale Gesetzgebung eine externe Prüfung durch einen ASAP vorschreibt (Krasodomska, Simnett & Street, 2021, S. 110). Der deutsche Gesetzgeber schreibt eine externe Prüfung hingegen nicht vor

[5] Die verwendeten Datenquellen sind auf dem beigefügten digitalen Datenträger unter dem Ordner „Stichprobe NBE“ einsehbar.

Tabelle 2: Stichprobenübersicht mit Angaben zum Unternehmenssitz und zur Branche (eigene Darstellung)

Unternehmen	Sitz	Branche
Aareal Bank AG	Wiesbaden	Finanzen
Aurubis AG	Hamburg	Energie & Rohstoffe
Brenntag AG	Essen	Maschinenbau, Bauwesen, Verkehr & Logistik
Ceconomy AG	Düsseldorf	Handel & Konsum
CTS Eventim AG & Co. KGaA	München	Medien
Delivery Hero AG	Berlin	Medien
Deutsche Wohnen SE	Berlin	Finanzen
Deutsche Pfandbriefbank AG	München	Finanzen
Evonik Industries AG	Essen	Chemie & Pharmazie
Fielmann AG	Hamburg	Handel & Konsum
Fraport AG	Frankfurt a. M.	Maschinenbau, Bauwesen, Verkehr & Logistik
FUCHS PETROLUB SE	Mannheim	Chemie & Pharmazie
GEA Group AG	Düsseldorf	Maschinenbau, Bauwesen, Verkehr & Logistik
Gerresheimer AG	Düsseldorf	Chemie & Pharmazie
Hannover Rück SE	Hannover	Finanzen
Hella GmbH & CO KGaA	Lippstadt	Maschinenbau, Bauwesen, Verkehr & Logistik
HOCHTIEF AG	Essen	Maschinenbau, Bauwesen, Verkehr & Logistik
HUGO BOSS AG	Metzingen	Handel & Konsum
innogy SE	Essen	Energie & Rohstoffe
Jungheinrich AG	Hamburg	Maschinenbau, Bauwesen, Verkehr & Logistik
K+S AG	Kassel	Chemie & Pharmazie
KION Group AG	Wiesbaden	Maschinenbau, Bauwesen, Verkehr & Logistik
LANXESS AG	Köln	Chemie & Pharmazie
LEG Immobilien AG	Düsseldorf	Finanzen
LEONI AG	Nürnberg	Maschinenbau, Bauwesen, Verkehr & Logistik
METRO AG	Düsseldorf	Handel & Konsum
MTU Aero Engines AG	München	Maschinenbau, Bauwesen, Verkehr & Logistik
NORMA Group SE	Maintal	Maschinenbau, Bauwesen, Verkehr & Logistik
Osram Licht AG	München	Maschinenbau, Bauwesen, Verkehr & Logistik
ProSiebenSat.1 Media SE	Unterföhring	Medien
PUMA SE	Herzogenaurach	Handel & Konsum
Rheinmetall AG	Düsseldorf	Maschinenbau, Bauwesen, Verkehr & Logistik
Rocket Internet	Berlin	Finanzen
Salzgitter AG	Salzgitter	Energie & Rohstoffe
Schaeffler AG	Herzogenaurach	Maschinenbau, Bauwesen, Verkehr & Logistik
Scout24 AG	München	Elektronik, Hard- und Software
Ströer SE & Co. KGaA	Köln	Medien
Symrise AG	Holzminden	Chemie & Pharmazie
TAG Immobilien AG	Hamburg	Finanzen
Talanx AG	Hannover	Finanzen
Zalando SE	Berlin	Handel & Konsum

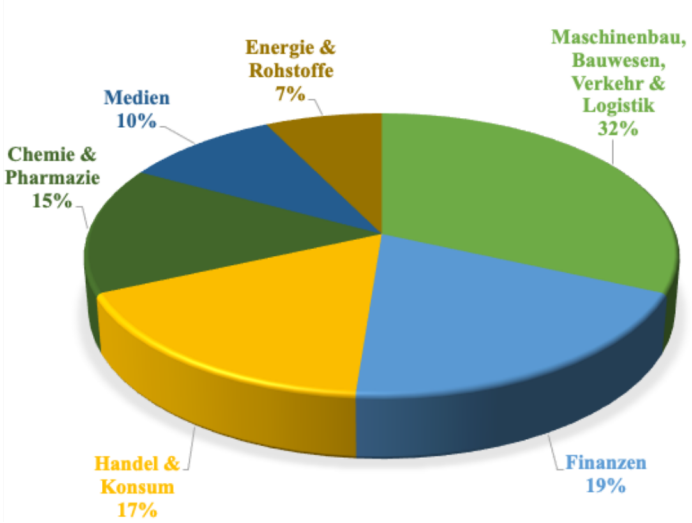


Abbildung 8: Branchenverteilung, N = 41 (eigene Darstellung)

Tabelle 3: Glaubwürdigkeitsindex - Standardisiertes Cronbach's Alpha, N = 41 (eigene Darstellung)

	W	WH	AV	GW
Vorher	0,819	0,875	0,740	0,917
Nachher	0,723	0,663	0,392	0,818

Tabelle 4: Glaubwürdigkeitsindex - Variablenübersicht (eigene Darstellung)

Subdimension	Variable	Variablenausprägung (Kodierung in Klammern)		
		WAHRHEIT		
Externe Prüfung	Anbieter	Nein (0)	ASAP (1)	NASAP (2)
	Prüfungsumfang	Selektiv (0)	Umfassend (1)	-
	Prüfungssicherheit	Mittel (0)	Hoch (1)	-
Eigenschaften der NBE	Verwendung eines Rahmenwerks	Nein (0)	GRI „Kern"/Andere (1)	GRI „Umfassend" (2)
	Seitenlänge der NFI	Seitenlänge: Normalisiert zwischen 0 und 1		
	Verortung der NFI	Integriert in Geschäftsber. (0)	Gesonderte NBE (1)	-
Genauigkeit	Methodik-Kapitel	Nein (0)	Ja (1)	-
	Genauigkeit der dargestellten Informationen	Durchschnittswert der Genauigkeit über drei Themen: Normalisiert zwischen 0 und 1		
		WAHRHAFTIGKEIT		
Wesentlichkeit	Existenz einer Wesentlichkeitsmatrix	Nein (0)	Ja (1)	-
	Aktualität der Wesentlichkeitsanalyse	Nein (0)	Ja (1)	-
	Bekenntnis des Vorstands zur Nachhaltigkeit	Nein (0)	Ja (1)	-
Stakeholderbeziehung	Identifikation	Nein (0)	Ja (1)	-
	Dialog	Nein (0)	Ja (1)	-
	Engagement	Nein (0)	Ja (1)	-
Organisationale Verankerung der Nachhaltigkeit	Organisationale Verankerung	Nein (0)	Ja (1)	Speziell für Nachhaltigkeit gegründet (2)
Sustainable Development Goals (SDGs)	Bezugnahme zu den SDGs	Nein (0)	Ja (1)	-
		ANGEMESSENHEIT & VERSTÄNDLICHKEIT		
Einfachheit in der Kommunikation	Lesbarkeit	FRE-Score: Normalisiert zwischen 0 und 1		
	Verwendung visueller Darstellungen	Anzahl visueller Seiten: Normalisiert zwischen 0 und 1		

(s. Kap. 3). Aus diesem Grund geht es für ein deutsches Unternehmen über das Mindestmaß gesetzlicher Vorschriften hinaus, wenn eine externe Prüfung stattfindet. Folglich ist die oben genannte Punkteverteilung gerechtfertigt.

Die Variable *Prüfungsumfang* wird folgt bestimmt: Sofern es sich bei der externen Prüfung um eine Prüfung handelt, bei der alle Inhalte der NBE geprüft werden, wird ein Punkt vergeben. Wird hingegen von der Prüfungsgesellschaft darauf hingewiesen, dass nur ausgewählte Sektionen der NBE geprüft wurden, erfolgt eine Bewertung mit der Punktzahl null.

Für die *Prüfungssicherheit* wird nur dann ein Punkt vergeben, wenn eine hohe Prüfungssicherheit vorliegt. Eine mittlere Prüfungssicherheit wird hingegen nicht mit einem Punkt honoriert.

Zur Identifikation der drei Variablen wird sowohl der Geschäftsbericht als auch - sofern vorhanden - der gesonderte Nachhaltigkeitsbericht auf einen Bestätigungsvermerk, der das Prüfungsurteil formal verbrieft, durch einen ASAP bzw. ein Zertifikat durch einen NASAP durchleuchtet. Dar-

in enthalten sind Angaben zur Prüfungsgesellschaft, dem Prüfungsumfang und der Prüfungssicherheit.

6.3.2. Eigenschaften der NBE (Wahrheit)

Falls das zu untersuchende Unternehmen vollständig auf die *Verwendung eines Rahmenwerks* für die NBE verzichtet, wird die Variable mit null Punkten bewertet. Bei der Orientierung an einem Rahmenwerk für die NBE (z.B. GRI Option „Kern“, DNK, Global Compact Index) wird ein Punkt vergeben. Im Unterschied zum Glaubwürdigkeitsindex nach Mazzotta et al. (2020), in dem alle betrachteten Unternehmen die GRI Standards befolgen, werden neben dem GRI Rahmenwerk auch andere Rahmenwerke mit einem Punkt bewertet. Grund hierfür ist, dass die Vorteile für die Verwendung eines Rahmenwerks nicht nur für das GRI Rahmenwerk zutreffend sind. Aufgrund der weitergehenden Anforderungen der GRI Option „Umfassend“ wird die Bewertung von zwei Punkten beibehalten.

Für die Bestimmung der Variable *Seitenlänge der NFI* wird approximativ die aggregierte Seitenlänge der NFI verwendet. Die Definition, was zu den NFI gehört, folgt den Ausführungen in Kapitel 2.2.[6] Demnach werden nur Informationen, welche den Säulen Umwelt und Soziales angehören, inkludiert. Des Weiteren werden Informationen zur Unternehmensführung, die nicht unmittelbar mit dem Nachhaltigkeitsmanagement in Verbindung stehen, nicht in der Seitenlänge berücksichtigt, da diese bereits vor dem CSR-RUG im § 289a HGB für die Unternehmen verpflichtend waren. Für das Aufsuchen der NFI wird der Geschäftsbericht und – sofern vorhanden – der Nachhaltigkeitsbericht verwendet. Falls es sich in den beiden Datenquellen um reine Wiederholungen, Kurzfassungen oder einen Verweis auf die gesonderte NBE handelt, wird dies nicht in der Variable berücksichtigt. Angebrochene Seiten von NFI werden als vollständige Seite gewertet. Die erlangten Werte werden normalisiert.[7] Aus der Normalisierung der Variable folgt, dass die minimale Seitenanzahl einen Punktwert von Null und die maximale Seitenanzahl einen Wert in Höhe von einem Punkt erhält. Die Definition der Variable weicht möglicherweise von der Herangehensweise Mazzotta et al. (2020, S. 1906) ab.

Die Variable *Verortung der NFI* wird mit einem Punkt versehen, wenn es eine gesonderte NBE gibt. Die im Kapitel 4.2) vorgestellten Variante a) (Integration der NFI im Lagebericht) als auch die Variante b) (Integrated Reporting) werden mit null Punkten versehen.

[6] Für eine genaue Ansicht der berücksichtigen NFI kann in der Excel-Datei „Datenerhebung“ unter dem Reiter „NFI im Sample“ betrachtet werden.

[7] Um die erlangten Werte sinnvoll in den Glaubwürdigkeitsindex einzubetten, muss eine Normalisierung der Variable *Seitenlänge der NFI* erfolgen. Die Formel zur Normalisierung für ein Unternehmen i lautet wie folgt: $Seitenlnge\ der\ NFI\ (normalisiert)_i = \frac{Seitenlnge\ der\ NFI_i - MIN}{MAX - MIN}$ (Helfaya & Whittington, 2019, S. 528). Mit dieser Vorgehensweise können metrische Skalen in eine metrische Skale, die Werte zwischen null und eins annimmt, umgewandelt werden. Für die Normalisierung werden die Daten vor und nach der CSR-RL betrachtet, damit ein sinnvoller ex-ante-ex-post-Vergleich möglich ist. Somit wird das globale Minimum bzw. Maximum (über beide Zeitpunkte hinweg) für die Berechnung verwendet.

6.3.3. Genauigkeit (Wahrheit)

Die Variable *Methodik-Kapitel* wird dann mit einem Punkt bewertet, wenn

- Angaben über die enthaltenen Entitäten gemacht werden,
- etwaige Änderungen in der Berichterstattung oder der Neudarstellung von Informationen transparent dargelegt werden,
- der Berichtszeitraum eindeutig definiert ist und
- das Vorgehen zur Bestimmung der Inhalte für die NBE erläutert wird.

Falls ein Unternehmen nicht alle Aspekte erwähnt, wird die Variable mit null Punkten angesetzt. Sofern ein Unternehmen GRI Rahmenwerk für die NBE verwendet, kann der GRI Content Index zur Vereinfachung der Inhaltsanalyse verwendet werden. Ansonsten werden die Aspekte händisch im Geschäfts- bzw. Nachhaltigkeitsbericht gesucht.

Die Variable *Genauigkeit der dargestellten Informationen* wird folgendermaßen bestimmt: Zuerst werden die wesentlichen Themen eines Unternehmens für beide Jahre identifiziert. Im besten Falle gibt es drei vergleichbare Themengebiete, die dann für beide Jahre hinsichtlich ihrer Genauigkeit untersucht werden. Ein Unternehmen erhält für ein Thema nur einen Punkt, wenn es sich um rein qualitative Inhalte handelt. Falls in dem Thema neben den qualitativen Informationen auch quantitative oder monetäre enthalten sind, werden zwei Punkte vergeben. In dem Fall, dass ein Thema qualitative, quantitative sowie monetäre Angaben vereint, werden drei Punkte für dieses Thema distribuiert. Die Punktzahl null wird nur dann für ein Thema verteilt, wenn es weniger als drei Themen gibt, über die berichtet wird. Nachdem alle drei Themen bewertet werden, wird ein Durchschnittswert über diese drei Themen gebildet. Diese Herangehensweise zur Bestimmung der Genauigkeit erfolgt in Anlehnung an wissenschaftliche Praxis (Helfaya & Whittington, 2019, S. 530; Michelon et al., 2015, S. 10). Falls der Durchschnittswert kleiner zwei ist, werden null Punkte vergeben. Durchschnittswerte, die größer gleich zwei sind, werden zwischen null und eins normalisiert.[8] Die konkrete Kodierung nach Mazzotta et al. (2020, S. 1903, 1906) weicht ggf. aufgrund fehlender Erläuterungen ab.

6.3.4. Wesentlichkeit (Wahrhaftigkeit)

Die Variable *Existenz einer Wesentlichkeitsmatrix* wurde dann mit einem Punkt versehen, wenn tatsächlich eine Matrix verwendet wurde (s. Kap. 2.5). Es muss folglich eine Priorisierung der Themen vorgenommen werden, die die Bedeutung des Themas für das Unternehmen und die Bedeutung für die Stakeholder widerspiegelt. Um den Punkt zu erreichen, ist es nicht nötig, eine im Vergleich zum Vorjahr veränderte Wesentlichkeitsmatrix zu präsentieren. Dieser Aspekt

[8] Die Formel für die Normalisierung lautet wie folgt: $Genauigkeit\ (normalisiert) = \frac{\phi Genawigkeit\ der\ drei\ betrachreten\ Themen_i - MIN}{MAX - MIN}$.

wird in der Variable *Aktualität der Wesentlichkeitsanalyse* aufgegriffen.

Für die Variable *Aktualität der Wesentlichkeitsanalyse* wird dann ein Punkt fällig, wenn neue bzw. veränderte wesentliche Themen ersichtlich sind oder explizit darauf hingewiesen wird, dass die Wesentlichkeitsanalyse neu geprüft worden ist. Im Unterschied zu Mazzotta et al. (2020, S. 1906) muss für den Erhalt eines Punktes keine Wesentlichkeitsmatrix existent sein. Bloße Auflistungen oder andere Darstellungen reichen bereits aus. Diese Bewertung resultiert aus dem Fakt, dass ein Großteil der Unternehmen keine Wesentlichkeitsmatrix verwenden (s. Kap. 6.5.1). Da diese Unternehmen bereits keinen Punkt für die Wesentlichkeitsmatrix erhalten, wird auf eine Doppelbestrafung verzichtet.

Die Variable *Bekenntnis des Vorstands zur Nachhaltigkeit* wird eins, wenn das Top-Management (i.d.R. vertreten durch den CEO) sich klar zur Verfolgung von Nachhaltigkeitszielen bereit erklärt. Hierfür wird primär darauf geachtet, ob im „CEO Letter" vom Thema Nachhaltigkeit hervorgehoben wird. Bei fehlendem Bekenntnis durch den Vorstand wird kein Punkt vergeben. Mazzotta et al. (2020, S. 1906) bezieht diese Variable auf die Wesentlichkeitsmatrix. Analog zum obigen Argument wird von dieser Kodierung abgewichen. Des Weiteren wird dieses Vorgehen in der Inhaltsanalyse von Mion und Loza Adaui (2019, S. 11) in dieser Weise implementiert.

6.3.5. Stakeholderbeziehung (Wahrhaftigkeit)

Die Variable *Identifikation* wird dann eins, wenn das Unternehmen eine Auflistung der verschiedenen Stakeholdergruppen vornimmt. Die Variablen *Dialog* und *Engagement* werden wie folgt bewertet: Im Einklang mit dem im Kapitel 2.5 beschriebenen unterschiedlichen Partizipationsgrad der Einbindung von Stakeholdern, wird für die Variable *Dialog* dann ein Punkt vergeben, wenn ein zweiseitiger Austausch zwischen dem Unternehmen und seinen Stakeholdern stattfindet (z.B. Etablierung von Diskussionsforen). Für die Variable *Engagement* wird nur dann ein Punkt vergeben, wenn die Stakeholder tatsächlich einen Einfluss auf die Entscheidungen der Unternehmung ausüben. Ein Beispiel hierfür kann das aktive Umsetzen von Vorschlägen seitens der Stakeholder (z.B. aus Diskussionen) sein oder eine Interessensvertretung im Aufsichtsrat des Unternehmens.

6.3.6. Organisationale Verankerung der Nachhaltigkeit (Wahrhaftigkeit)

Bei der Variable *Organisationale Verankerung* wird kein Punkt vergeben, wenn aus den Informationen der NBE nicht hervorgeht, dass die Nachhaltigkeit organisational im Unternehmen eingegliedert ist. Außerdem wird kein Punkt vergeben, wenn darauf hingewiesen wird, dass der Vorstand bzw. der Aufsichtsrat die alleinige Verantwortung trägt. Da dies faktisch durch die rechtliche Stellung beider Organe determiniert ist, ist dieser Verweis nicht ausreichend, damit ein Punkt vergeben werden kann. Falls das Thema Nachhaltigkeit in eine bestehende Organisation bzw. Abteilung integriert ist (z.B. Abteilung Compliance), dann erhält das Unternehmen einen Punkt. Speziell für das Thema Nachhaltigkeit gegründete Strukturen (z.B. Sustainability Committee) werden mit zwei Punkten bewertet.

6.3.7. Sustainable Development Goals (Wahrhaftigkeit)

Die Variable *Bezugnahme zu den SDGs* wird dann eins, wenn das betrachtete Unternehmen in irgendeiner Weise die SDGs referenziert. Sofern die SDGs nicht adressiert werden, ist die Variable gleich null.

6.3.8. Einfachheit in der Kommunikation (Angemessenheit & Verständlichkeit)

Die Subdimension Einfachheit in der Kommunikation besteht aus den Variablen *Lesbarkeit* und *Verwendung visueller Darstellungen*. Für beide Variablen wird eine zufällige Stichprobe von zehn Seiten mittels Excel-Funktion „Zufallsbereich" gezogen.[9] Der Lesbarkeitsindex Flesch-Reading-Ease (FRE) hilft dabei, die Variable *Lesbarkeit* zu quantifizieren. Der FRE-Score kann Werte zwischen null und 100 annehmen. Ein höherer FRE-Score steht für eine bessere Lesbarkeit des Textes (Flesch, 1948, S. 230). Die von Rudolf Flesh entwickelte Formel lautet folgendermaßen: $FRE-Score = 206,835 - (1,015 * \phi Satzlnge) - (84,6 * \phi Silbenanzahl proWort)$ (Flesch, 1948, S. 225).[10] Wenn die durchschnittliche Satzlänge oder die durchschnittliche Silbenanzahl pro Wort steigt, sinkt der FRE-Score. Anders ausgedrückt heißt dies, dass die Lesbarkeit mit einer höheren durchschnittlichen Satzlänge sowie steigender durchschnittlicher Silbenanzahl pro Wort abnimmt. Demnach erhalten jene Unternehmen einen besseren FRE-Score, wenn kurze Sätze und Worte verwendet werden. Bei Unternehmen, welche weniger als zehn auswertbare Seiten NFI haben, werden alle vorhandenen Seiten mit NFI in die Zufallsstichprobe einbezogen. Für die Berechnung des FRE-Scores, wird die browserbasierte Open Source Software von „leichtlesbar.ch" benutzt. Nachdem ein Word-Dokument aufbereitet wird, das die gezogene Zufallsstichprobe beinhaltet, wird per Copy-and-Paste-Befehl der Text aus der NBE in die Eingabemaske der Software eingespielt. Die FRE-Score-Werte werden die Werte zwischen null und eins normalisiert.[11] Die Autoren werten die Lesbarkeit Mazzotta et al. (2020, S. 1906) anhand des Gulpease Index aus. Da dieser Lesbarkeitsindex speziell für die italienische Sprache entwickelt worden ist, kann dieser nicht zur Bestimmung der Lesbarkeit verwendet werden. Aus diesem Grund wurde auf den Lesbarkeitsindex von Flesch, der für die englische Sprache ausgerichtet ist, ausgewichen.

Die Variable *Verwendung visueller Darstellungen* wird mit folgender Kennzahl berechnet: $Verwendung\ visueller$

[9] Die Zufallsstichproben sind im Ordner „Stichprobe (NBE)" für jedes Unternehmens unter „Kommunikation" einsehbar.

[10] $\phi Satzlaenge = \frac{Anzahl\ der\ Woerter\ im\ Text}{Anzahl\ der\ Saetze\ im\ Text}$, $\phi\ Silbenanzahl\ pro\ Wort = \frac{Silbenanzahl\ im\ Text}{Anzahl\ der\ Woerter\ im\ Text}$

[11] Für die Normalisierung wird folgende Formel verwendet: $FRE - Score\ (normalisiert)_i = \frac{FRE-Score_i-MIN}{MAX-MIN}$.

$Darstellungen_i = \frac{\sum_{i=1}^{10} w_j * x_{ij}}{10}$.[12] Eine Seite wird dann als visuelle Seite eingestuft, wenn mindestens eine Abbildung, eine Tabelle oder ein Bild zu erkennen ist. Sofern es weniger als zehn Seiten NFI gibt, werden alle Seiten untersucht. Da die Stichprobengröße der gezogenen Seiten dann unter zehn Seiten liegt, wird entsprechend Nenner auf die Anzahl der betrachteten Seiten adjustiert. Um die obige Kennzahl in die Variable Anzahl visueller Seiten zu transformieren, erfolgt eine Normalisierung der Werte zwischen null und eins.[13] Mazzotta et al. (2020) machen keine Angaben darüber, ob für jedes Unternehmen alle Seiten untersucht wurden, oder ob man sich auf eine Stichprobe reduziert hat. Folglich ist die Vorgehensweise u.U. divergent.

6.4. Messung der Kontrollvariablen

In der bisherigen Literatur gibt es zahlreiche empirische Befunde über Determinanten, die die Qualität der NBE positiv wie auch negativ beeinflussen können (Ali et al., 2017, S. 276-287; Dienes et al., 2016, S. 167). Die Qualität der NBE wird in der Literatur mit verschiedenen Begriffen wie „sustainability reporting quality“, „environmental disclosure quality“, etc. beschrieben (Cormier, Magnan & Van Velthoven, 2005, S. 3; Mion & Loza Adaui, 2019, S. 11). Wenngleich nicht explizit von Glaubwürdigkeit der NBE die Rede ist, handelt es sich bei der Glaubwürdigkeit um ein Teilaspekt der Qualität der NBE (Helfaya et al., 2018, S. 19; Mion & Loza Adaui, 2019, S. 12). Demnach steht dem Übertragen der Kontrollvariablen wie im wissenschaftlichen Artikel von Mion und Loza Adaui (2019, S. 12) in den Kontext der Glaubwürdigkeit der NBE nichts im Wege. Im Folgenden wird auf insgesamt sechs Kontrollvariablen eingegangen, die neben der CSR-RL eine Veränderung der Glaubwürdigkeit in der NBE verursachen können. In Tabelle 5 ist eine Übersicht der Kontrollvariablen illustriert. Die Daten für die Kontrollvariablen wurden anhand der Geschäfts- bzw. Nachhaltigkeitsberichte oder dem Bundesanzeiger entnommen.

6.4.1. Unternehmensgröße

Grundsätzlich wird in der Betriebswirtschaftslehre davon ausgegangen, dass es zwischen dem Management und den Investoren (z.B. Aktionäre, Fremdkapitalgeber) eine Informationsasymmetrie gibt (Cormier, Ledoux, Magnan & Aerts, 2010, S. 574). Im Kontext der NBE sind die Investoren neben finanziellen Informationen auch an NFI interessiert (Amel-Zadeh & Serafeim, 2018, S. 87). Sofern das Management eine unglaubwürdige NBE bzw. gänzlich auf die Veröffentlichung von NFI verzichtet, müssen die einzelnen Investoren in eigener Regie Informationen sammeln. Aus Investorensicht ist dieser Informationsbeschaffungsprozess mit enormen Informationskosten verbunden (Cormier et al., 2005, S. 9). Die Informationskosten sind für die Investoren geringer, wenn das Unternehmen freiwillig NFI offenlegt. Eine freiwillige Veröffentlichung von NFI ergibt aus der Unternehmenssicht nur dann Sinn, wenn der Nutzen aus der NBE höher ist als die daraus resultierenden Kosten (Erstellung des Berichts, Druckkosten, etc.) (Cormier et al., 2005, S. 9; M. Fifka, 2014, S. 10-11). Für ein Unternehmen, das bspw. stark auf den Kapitalmarkt angewiesen ist, sinken die Finanzierungskosten durch die Offenlegung von NFI. Dies kann mit der sinkenden Informationsasymmetrie und dem damit geringeren Risiko für Investoren begründet werden (Cormier et al., 2005, S. 9; Frankel, McNichols & Wilson, 1995, S. 135). Somit kann der Nutzen aus der Offenlegung von NFI die Informationskosten überwiegen. Die Informationskosten für ein großes Unternehmen sind relativ zu den Informationskosten eines kleinen Unternehmens geringer (Cormier & Magnan, 1999, S. 439). Demnach ist es wahrscheinlicher, dass ein großes Unternehmen mehr NFI als ein kleines Unternehmen publiziert, da der Nutzen aus der Offenlegung der NFI die Informationskosten eher kompensiert. Da der Nutzen aus der Offenlegung von NFI in einem großen Unternehmen die Informationskosten eher kompensiert, ist von einer positiven Wirkung auf die Glaubwürdigkeit in der NBE auszugehen.

Die Unternehmensgröße ist definiert als der natürliche Logarithmus der Bilanzsumme und erfolgt in Anlehnung an zahlreiche wissenschaftliche Publikationen (Dienes et al., 2016, S. 167).

6.4.2. Eigentumsstreuung

Ein Unternehmen kann eine geringe Eigentumsstreuung („concentrated ownership“) oder hohe Eigentumsstreuung („dispersed ownership“) aufweisen (Brammer & Pavelin, 2008, S. 124). Das Konfliktpotential zwischen dem Management und den Eigentümern ist bei einer hohen Eigentumsstreuung eher gegeben (Gamerschlag, Möller & Verbeeten, 2011, S. 238). Zum einen haben die Anteilseigner weniger Kontrolle über das Management, weshalb die Offenlegung von NFI zu einem wichtigen Kommunikationsmedium wird (Berthelot, Coulmont & Serret, 2012, S. 361; Brammer & Pavelin, 2006, S. 1173). Dies ermöglicht, dass die Eigentümer regelmäßig über den Status quo informiert werden und im Sinne der Eigentümer agiert wird (Chau & Gray, 2002, S. 249). Bei einer geringen Eigentumsstreuung, wie z.B. in einem familiengeführten Unternehmen, ist die Nachfrage nach NFI geringer. Folglich ist die Motivation, eine qualitativ hochwertige und über das Mindestmaß der Gesetzgebung hinausgehende NBE zu erstellen, gering (Chau & Gray, 2002, S. 250, 258). Basierend auf den vorherigen Ausführungen ist von einer glaubwürdigeren NBE bei einer hohen Eigentumsstreuung auszugehen, da die Informationsasymmetrie bei einer hohen Eigentumsstreuung größer ist und somit ist die Nachfrage der zahlreichen Stakeholder nach NFI stärker ausgeprägt.

In Anlehnung an Cormier und Magnan (2003, S. 50) wird eine binäre Dummy-Variable eingeführt, um die Eigentümerstruktur zu berücksichtigen. Diese ist gleich eins, wenn es kei-

[12] Die Abkürzungen sind folgendermaßen zu interpretieren: $w_j = 1$, wenn die Seite j visuell ist, $w_j = 0$, sonst; x_{ij} = Seite j von Unternehmen i

[13] Für die Normalisierung wird folgende Formel verwendet: $Verwendung\ visueller\ Darstellungen\ (normalisiert)_i = \frac{Verwendung\ pisueller\ Darsteliungen_i - MIN}{MAX - MIN}$

Tabelle 5: Kontrollvariablen - Variablenübersicht (eigene Darstellung)

Kontrollvariable	Kürzel	Definition
Unternehmensgröße	SIZE	ln(Bilanzsumme) = ln(Eigenkapital + Fremdkapital)
Eigentumsstreuung	DISP	$DISP = \begin{cases} 1, kein\ Antetlsetgnter\ mit\ mehr\ als\ 20\%\ der\ Stimmrechte \\ 0, sonst \end{cases}$
Ertragslage	ROA	Gesamtkapitalrendite = ((Gewinn + Fremdkapitalzinsen)/Gesamtkapital)*100 %
Vermögenslage	LEV	Verschuldungsgrad = (Fremdkapital/Eigenkapital)*100 %
Medienpräsenz	PRES	# Suchtreffer im Handelsblatt-Archiv für das relevante Geschäftsjahr
Umweltsensitivität	ENVSEN	$ENVSEN = \begin{cases} 1, hohe\ Umweitsensitivitatt\ liegt\ vor \\ 0, sonst \end{cases}$

nen Anteilseigner gibt, der mehr als 20 %[14] der Stimmrechte innehat.

6.4.3. Vermögens- und Ertragslage

Eine vollkommen transparente Offenlegung von NFI kann nachteilig für das Unternehmen sein. Die Kommunikation interner Informationen in die Öffentlichkeit kann dazu führen, dass Dritte diese Informationen gegen das Unternehmen verwenden. Dies kann Unternehmen dazu verleiten, eine selektive NBE zu betreiben, damit sich keine Auswirkungen auf den Gewinn ergeben (Cormier et al., 2005, S. 9). Stabile wirtschaftliche Verhältnisse helfen dabei, Kritik, die potentiell aus den veröffentlichten NFI droht, auszuhalten (Cormier et al., 2005, S. 10-11). Ferner unterliegen profitable Firmen eher einem Rechtfertigungsdruck bezogen auf ihren gesellschaftlichen Mehrwert als weniger profitable Firmen (Bewley & Li, 2000, S. 208-209). Ein weiteres Argument ist, dass profitable Firmen tendenziell eher bereit sind, die Gewinne in eine qualitativ hochwertige NBE zu reinvestieren (Brammer & Pavelin, 2006, S. 1174). Es bleibt somit festzuhalten, dass stabile wirtschaftliche Verhältnisse einer qualitativ hochwertigen NBE und damit der Glaubwürdigkeit dienlich sind.

Um die Ertragslage bzw. die Profitabilität zu quantifizieren, wird die Gesamtkapitalrendite in % berechnet (z.B. Mion & Loza Adaui, 2019, S. 16). Für die Vermögenslage wird der Verschuldungsgrad in % verwendet (z.B. Cormier et al., 2005, S. 18).

6.4.4. Medienpräsenz

In der öffentlichen Wahrnehmung gibt es erhebliche Unterschiede zwischen einzelnen Unternehmen. Unternehmen, die regelmäßig in den Medien präsent sind, unterliegen einem steten gesellschaftlichen und politischen Druck (Brammer & Millington, 2006, S. 6-7). Aus diesem Grund wird für die Sichtbarkeit in den Medien kontrolliert. Unternehmen, die ständig in den Medien zu sehen sind, könnten demnach eine bessere NBE aufweisen als Unternehmen, die weniger mediale Aufmerksamkeit erhalten. Eine hohe Medienpräsenz wirkt sich folglich positiv auf die Qualität der NBE und damit der Glaubwürdigkeit aus.

Hierfür werden die Anzahl der Suchtreffer im Archiv des Handelsblatts für das relevante Geschäftsjahr betrachtet. Das Handelsblatt wurde ausgewählt, weil es sich um eine renommierte Wirtschaftszeitung mit einer hohen Reichweite in Deutschland handelt (Gamerschlag et al., 2011, S. 243-244).

6.4.5. Umweltsensitivität

Ein Chemieunternehmen muss naturgemäß der NBE einen höheren Stellenwert einräumen als ein IT-Unternehmen oder ein Finanzdienstleister (Helfaya & Whittington, 2019, S. 531; Meek, Roberts & Gray, 1995, S. 559). Branchen unterscheiden sich erheblich bzgl. ihres Einflusses auf die Umwelt. Umweltsensitive Unternehmen sind eher im Fokus von Umwelt- und Naturschutzverbänden, die wiederum die öffentliche Wahrnehmung und die Politik beeinflussen (Deegan & Gordon, 1996, S. 191). Aus diesem Grund kann eine bessere Qualität der NBE und damit der Glaubwürdigkeit bei Unternehmen erwartet werden, die eine hohe Umweltsensitivität besitzen.

Um diese branchenspezifischen Unterschiede zu berücksichtigen, werden die unterschiedlichen Branchen (s. Kap. 6.2) in folgende drei Gruppen nach Helfaya und Whittington (2019, S. 531) eingeteilt: (a) hohe Umweltsensitivität, (b) mittlere Umweltsensitivität sowie eine (c) geringe Umweltsensitivität. Es wird eine Dummy-Variable eingeführt, die den Wert eins annimmt, wenn eine hohe Umweltsensitivität vorliegt. Im Anhang in Tabelle 12 ist dargelegt, welches Unternehmen zu welcher Umweltsensitivitätsgruppe gehört.

6.5. Ergebnisse

Das Kapitel 6.5 unterteilt sich in drei Unterkapitel. Kapitel 6.5.1 beschreibt die Resultate der Inhaltsanalyse deskriptiv, während Kapitel 6.5.2 auf die induktiven Ergebnisse eingeht. Im Kapitel 6.5.3 werden die Befunddaten diskutiert. Zur statistischen Auswertung werden die Statistikprogramme SPSS sowie Gretl verwendet.[15]

[14] Gemäß der International Accounting Standards 28 (2011) § 5 ist ab einem Stimmrechtsanteil von 20 % ein signifikanter Einfluss auf das Unternehmen möglich.

[15] Abgesehen von den im Kapitel 6.5 aufgeführten Ergebnissen können auf dem beigefügten digitalen Datenträger weitergehende statistische Auswertungen (z.B. die Prüfung der Voraussetzungen zur Durchführung einer Regressionsanalyse) in den Ordnern „Deskriptive Statistik/Nichparametrisch(SPSS)“ sowie „Parametrisch(Gretl)“ eingesehen werden.

6.5.1. Deskriptive Ergebnisse

In Tabelle 6 sind die Ergebnisse der Inhaltsanalyse dargestellt. Für eine umfassendere Perspektive werden die deskriptiven Ergebnisse auf aggregierter Ebene (Spalte Gesamt in Tab. 6) sowie auf Branchenebene in den nachfolgenden Ausführungen dargestellt. Insbesondere wird dabei auf die relative Veränderung eingegangen (und nicht auf das absolute Niveau), damit deutlich wird, ob es sich um eine Verschlechterung oder Verbesserung nach der Einführung der CSR-RL handelt.[16]

Externe Prüfung (Wahrheit)

Die Subdimension externe Prüfung unterteilt sich in die Variablen *Anbieter*, *Prüfungsumfang* sowie *Prüfungssicherheit*.

Es zeigt sich, dass der Anteil der Unternehmen ohne externe Prüfung von 73 % auf 30 % gesunken ist. Alle bis auf ein Unternehmen (2 %) entschieden sich dazu, die externe Prüfung durch einen ASAP durchführen zu lassen. Alle umgesetzten externen Prüfungen wurden durch einen ASAP, nämlich der Big-Four-Wirtschaftsprüfungsgesellschaften (PwC, Deloitte, EY oder KPMG), durchgeführt. Die Symrise AG beauftragte die DQS CFS GmbH als einzigen NASAP zum externen Prüfungsdienstleister. Auffällig ist, dass alle Branchen mit Ausnahme der Chemie- und Pharmazieunternehmen eine erhöhte Bereitschaft für eine externe Prüfung aufzeigen (Anstiege zwischen +28 bis 77 %-Punkte). Die Branche Chemie & Pharmazie zeigte keine Veränderung.

Nur ein geringer Anteil von 5 % der Unternehmen hat sich vor der CSR-RL für eine umfassende externe Prüfung entschieden. Der Anteil ist nach Einführung der CSR-RL auf 32 % gestiegen. Die meisten Branchen weisen eine Verbesserung hinsichtlich des Prüfungsumfangs auf (Anstiege zwischen +25 bis 50 %-Punkte). Erneut verbessert sich die Branche Chemie & Pharmazie nicht nach der Einführung der CSR-RL.

100 % der Unternehmen entschieden sich wenn überhaupt für eine mittlere Prüfungssicherheit. Eine Unterschiedsbetrachtung zwischen den Branchen entfällt folglich.

Eigenschaften der NBE (Wahrheit)

Die Variablen *Verwendung eines Rahmenwerks*, *Seitenlänge der NFI* sowie die *Verortung der NFI*, gehören zu der Subdimension Eigenschaften der NBE.

Anhand der Tabelle 6 ist zu erkennen, dass der Anteil der Unternehmen, die ein externes Rahmenwerk für die NBE verwenden, mit einem 20 %-igen Zuwachs deutlich gestiegen ist. Vor der CSR-RL benutzten 56 % GRI „Kern“ und 2 % GRI „Umfassend“. Die übrigen 42 % wenden kein Rahmenwerk[17] für die NBE an. Nach der CSR-RL sieht die Verteilung wie folgt aus: 76 % der Unternehmen nutzen GRI Option „Kern“ oder andere Rahmenwerke (DNK, Global Compact Index) und 2 % GRI Option „Umfassend“. Die verbleibenden 22 % der Unternehmen nutzen kein Rahmenwerk für ihre NBE. Die Symrise AG ist das einzige Unternehmen mit diesem Anwendungsniveau. Während bei den Branchen Chemie & Pharmazie sowie Medien keine Änderung stattfindet, verbuchen alle anderen Branchen Zuwächse (zwischen +12 bis 77 %-Punkte) in der Anwendung von Rahmenwerken.

Vor der CSR-RL beträgt die globale minimale Seitenzahl der betrachteten NFI null und die globale maximale Seitenlänge 210. Die Berechnungsformel zur Ermittlung der normalisierten Seitenlänge der NFI für Unternehmen i lautet demnach: $Seitenlnge\ der\ NFI\ (normalisiert)_i = \frac{Seitenlange\ der\ NFI}{210}$.[18] Insgesamt ist die Länge der NFI von 0,21 vor der CSR-RL verglichen mit 0,30 nach der CSR-RL um ca. 43 % gestiegen. Während es bei den Branchen Medien (+100 %) Die Branche Handelund Handel & Konsum (+108 %) zu einer Verdopplung der NFI kommt, weisen die anderen Branchen moderatere Anstiege auf (Chemie & Pharmazie: +10 %, Maschinenbau, Bauwesen, Verkehr & Logistik: +24 %, Finanzen: +48 %, Energie & Rohstoffe: +58 %).

Der Anteil der Unternehmen, die ihre NFI in einer gesonderten NBE veröffentlichen, erhöht sich um 24 %-Punkte nach Inkrafttreten der CSR-RL (von 49 % auf 73 %). Markant an dieser Stelle ist, dass die Branchen Finanzen, Medien sowie Energie & Rohstoffe nach der CSR-RL zu 100 % eine gesonderte NBE aufweisen. Während die Chemie & Pharmazie Branche (-17 %-Punkte) die Möglichkeit einer gesonderten NBE weniger nach der CSR-RL nutzt, weisen die Branchen Maschinenbau, Bauwesen, Verkehr & Logistik (+8 %-Punkte) sowie Handel & Konsum (+14 %-Punkte) einen gemäßigten Anstieg auf.

Genauigkeit (Wahrheit)

Die Subdimension Genauigkeit untergliedert sich in die Variablen *Methodik-Kapitel* und *Genauigkeit der dargestellten Informationen*.

Insgesamt steigt die Verwendung eines Methodik-Kapitels nach der CSR-RL um 22 %-Punkte von 56 % auf 78 %. In der Branche Chemie & Pharmazie ergeben sich keine Änderungen zum Vorjahr und die übrigen Branchen verzeichnen Zuwächse zwischen 12 bis 77 %-Punkten.

Die Genauigkeit der dargestellten Informationen stieg bei aggregierter Betrachtungsweise um 40 % von 0,25 auf 0,35. Das globale Minimum beträgt zwei (Durchschnittswerte unter zwei werden unmittelbar mit null Punkten versehen) und das globale Maximum ist drei. Daraus vereinfacht sich Berechnungsformel für die normalisierte Genauigkeit für Unternehmen i wie folgt: $Genauigkeit(normalisiert)_i = \phi Genauigkeit_i - 2$.[19] Die Branche Handel & Konsum liefert

[16] Die Inhaltsanalyse auf Unternehmensebene kann in der Excel-Datei „Datenerhebung“ in den Reitern „Daten nach CSR-RL“ sowie „Daten vor CSR-RL“ eingesehen werden. Die Aufbereitung der Daten erfolgte in der Datei „Datenerhebung“.

[17] Unternehmen in der Stichprobe, die nur ein Rahmenwerk referenzieren, aber es nicht tatsächlich anwenden (d.h. nicht mindestens GRI Option „Kern“ vorliegt) sind in „kein Rahmenwerk“ inkludiert (Mazzotta et al., 2020, S. 1907).

[18] Hinweis: Mit Hilfe dieser Formel kann ausgehend von der normalisierten Seitenlänge aus Tabelle 6 die durchschnittliche Seitenanzahl von NFI berechnet werden: Seitenlänge (normalisiert) x 210 = Durchschnittliche Seitenlänge (gesamt bzw. nach Branche).

[19] Beispiel für die Interpretation: Die durchschnittliche Genauigkeit liegt

Tabelle 6: Deskriptive Statistik auf Variablenebene und nach Branche, N = 41 (eigene Darstellung)

Dimension/ Subdimension/ Variable	Variablenausprägung	Maschinenbau, Bauwesen, Verkehr & Logistik		Finanzen		Handel & Konsum		Chemie & Pharmazie		Medien		Energie & Rohstoffe		Gesamt	
	N	13		8		7		6		4		3		41	
	CSR-RL	vorher	nachher	vorher	nachher	vorher	nachher	vorher	nachher	vorher	nachher	vorher	nachher	vorher	nachher
WAHRHEIT															
Externe Prüfung															
Anbieter	Keine	85 %	23 %	88 %	38 %	57 %	29 %	50%	50 %	75 %	25 %	67 %	0 %	73 %	30 %
	ASAP	15 %	77 %	12 %	62 %	43 %	71 %	33 %	33 %	25 %	75 %	33 %	100 %	25 %	68 %
	NASAP	0 %	0 %	0 %	0 %	0 %	0 %	17 %	17 %	0 %	0 %	0 %	0 %	2 %	2 %
Prüfungsumfang	Selektiv oder keine Prüfung	100 %	69 %	100 %	50 %	86 %	50 %	83 %	83 %	100 %	75 %	100 %	67 %	95 %	68 %
	Umfassend	0 %	31 %	0 %	50 %	14 %	50 %	17 %	17 %	0 %	25 %	0 %	33 %	5 %	32 %
Prüfungssicherheit	Mittel oder keine Prüfung	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
	Hoch	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Eigenschaften der NBE															
Verwendung eines Rahmenwerks	Nein	23 %	8 %	50 %	25 %	57 %	29 %	33 %	33 %	50 %	50 %	67 %	0 %	42 %	22 %
	GRI „Kern“/Andere	77 %	92 %	50 %	75 %	43 %	71 %	50 %	50 %	50 %	50 %	33 %	100 %	56 %	76 %
	GRI „Umfassend“	0 %	0 %	0 %	0 %	0 %	0 %	17 %	17 %	0 %	0 %	0 %	0 %	2 %	2 %
Länge der NFI	Durchschnitt (normalisiert)	0,29	0,36	0,21	0,31	0,13	0,27	0,20	0,22	0,11	0,22	0,19	0,30	0,21	0,30
Verortung der NFI	Im AR integriert	31 %	23 %	50 %	0 %	71 %	57 %	50 %	67 %	75 %	0 %	67 %	0 %	51 %	27 %
	Gesonderte NBE	69 %	77 %	50 %	100 %	29 %	43 %	50 %	33 %	25 %	100 %	33 %	100 %	49 %	73 %
Genauigkeit der dargestellten Informationen															
Methodik-Kapitel	Nein	23 %	8 %	50 %	38 %	57 %	29 %	33 %	33 %	75 %	25 %	67 %	0 %	44 %	22 %
	Ja	77 %	92 %	50 %	62 %	43 %	71 %	67 %	67 %	25 %	75 %	33 %	100 %	56 %	78 %
Genauigkeit	Durchschnitt (normalisiert)	0,31	0,41	0,25	0,38	0,10	0,10	0,39	0,44	0,17	0,33	0,22	0,44	0,25	0,35
WAHRHAFTIGKEIT															
Wesentlichkeit															
Existenz einer Wesentlichkeitsmatrix	Nein	46 %	39 %	62 %	50 %	86 %	57 %	33 %	50 %	100 %	100 %	100 %	100 %	63 %	56 %
Organisationale Verankerung der Nachhaltigkeit															
Organisationale Verankerung	Nein	46 %	31 %	63 %	25 %	57 %	29 %	33 %	17 %	50 %	50 %	67 %	33 %	51 %	29 %
	Ja	0 %	0 %	12 %	25 %	0 %	0 %	0 %	0 %	25 %	25 %	0 %	0 %	5 %	7 %
	Speziell für Nachhaltigkeit	54 %	69 %	25 %	50 %	43 %	71 %	67 %	83 %	25 %	25 %	33 %	67 %	44 %	64 %
Sustainable Development Goals (SDGs)															
Bezugnahme zu den SDGs	Nein	54 %	39 %	88 %	63 %	57 %	43 %	50 %	50 %	100 %	100 %	100 %	33 %	68 %	51 %
	Ja	46 %	61 %	12 %	37 %	43 %	57 %	50 %	50 %	0 %	0 %	0 %	67 %	32 %	49 %
ANGEMESSENHEIT & VERSTÄNDLICHKEIT															
Einfachheit in der Kommunikation															
Lesbarkeit	Durchschnitt (normalisiert)	0,60	0,58	0,51	0,54	0,52	0,66	0,58	0,52	0,61	0,65	0,52	0,60	0,56	0,58
Verwendung visueller Darstellungen	Durchschnitt (normalisiert)	0,55	0,70	0,26	0,34	0,59	0,60	0,56	0,53	0,625	0,33	0,38	0,43	0,50	0,53

nach der CSR-RL keine genaueren Informationen und die übrigen Branchen zeigen moderate bis starke Zuwachsraten (Chemie & Pharmazie: +13%, Maschinenbau, Bauwesen, Verkehr & Logistik: +32 %, Finanzen: +52 %, Medien: +94 %).

Wesentlichkeit (Wahrhaftigkeit)

Die Subdimension Wesentlichkeit vereint die Variablen *Existenz einer Wesentlichkeitsmatrix*, *Aktualität der Wesentlichkeitsanalyse* sowie *Bekenntnis des Vorstands zur Nachhaltigkeit*.

Für die Bestimmung der wesentlichen Themen nutzen vor der CSR-RL 37 % der untersuchten Unternehmen eine Wesentlichkeitsmatrix zur Visualisierung. Nach der CSR-RL stieg der Anteil auf 44 % und damit um 7 %-Punkte. Sowohl vor als auch nach der CSR-RL verzichten die Branchen Medien und Energie & Rohstoffe gänzlich auf das Erstellen einer Wesentlichkeitsmatrix. Die Branche Chemie & Pharmazie verzeichnet einen Rückgang um 17 %-Punkte (von 67 % auf 50 %) und die restlichen Branchen zeigen Zuwächse von 7 bis 29 %-Punkten.

In der Gesamtschau steigt der Anteil der Unternehmen, welche die wesentlichen Themen aktualisiert haben, von 39 % auf 83 %. Bei allen Branchen sind Zuwächse zu erkennen,

nach der CSR-RL insgesamt bei 2,35. Die normalisierte Genauigkeit beträgt folglich 2,35 – 2 = 0,35.

welche zwischen 31 bis 75 %-Prozentpunkten liegen.

Ebenso steigt die Anzahl der Vorstände, die sich zur Nachhaltigkeit bekennen, von 58 % auf 80 %. Während alle Branchen einen höheren Anteil nach der CSR-RL aufweisen (Anstiege zwischen +12 bis 43 %-Punkte), ist für die Branche Energie & Rohstoffe keine Verbesserung zu erkennen.

Stakeholderbeziehung (Wahrhaftigkeit)

Die Subdimension Stakeholderbeziehung spaltet sich in die Variablen *Identifikation*, *Dialog* und *Engagement* auf.

Nach der CSR-RL schaffen es 88 % der Unternehmen in ihrer NBE, relevante Stakeholdergruppen zu nennen. Damit liegt ein Wachstum von 30 %-Punkten vor im Vergleich zum Vorjahr vor. Hervorzuheben sind die Branchen Energie & Rohstoffe (+77 %-Punkte) und Handel & Konsum (57 %-Punkte) mit deutlichen Anteilszuwächsen. Die übrigen Branchen weisen ein mittleres Anstiegsniveau zwischen 15 bis 26 %-Punkten auf.

Daneben findet nach der CSR-RL häufiger ein Dialog mit den Stakeholdern statt (Anstieg von 61 % auf 85 %). Während das Niveau bei Handel & Konsum (71 %) sowie Chemie & Pharmazie (83 %) auf einem konstant hohen Niveau bleibt, steigern sich die übrigen Branchen bedeutsam. Die Branchen Maschinenbau, Bauwesen, Verkehr & Logistik sowie Energie & Rohstoffe erreichen sogar die 100 %-Marke.

Während sich für die ersten beiden Variablen *Identifikation* und *Dialog* eine positive Entwicklung nach der Gesetzesreform abzeichnet, gilt dies nicht für die letzte Variable der Subdimension. Die Variable *Engagement* verweilt auf einem konstant niedrigen Level von 17 %. Das heißt, dass 83 % der Unternehmen ihre Stakeholder nicht aktiv in den Entscheidungsprozess involvieren. Im relativen Vergleich der Branchen schneiden die Branchen Chemie & Pharmazie sowie Energie & Rohstoffe mit einem Anteil über 33 % in beiden Geschäftsjahren am besten ab.

Organisationale Verankerung der Nachhaltigkeit (Wahrhaftigkeit)

Die Subdimension Organisationale Verankerung der Nachhaltigkeit besteht ausschließlich aus der Variable *Organisationale Verankerung*. Insgesamt zeigen vor der CSR-RL 49 % der Unternehmen eine organisationale Verankerung. In dem Geschäftsjahr danach sind es bereits 71 %. Nur in seltenen Fällen (5 % vor der CSR-RL, 7 % danach) existiert eine organisationale Struktur, in dem nachhaltigkeitsbezogene Themen in eine bestehende Organisation eingegliedert werden. Nur in den Branchen Finanzen und Medien wird die Nachhaltigkeit in bereits vorhandene organisationale Strukturen integriert. Die Branche Medien zeigt keine Variation vor und nach der CSR-RL. Hingegen steigert sich der Anteil der Unternehmen mit einer organisationalen Verankerung in der Branche Finanzen von 37 % auf 75 % merklich. Bei den verbleibenden Unternehmen ist überall ein Anteilszuwachs ersichtlich. Dieser liegt zwischen 15 % und 34 %-Punkten.

Sustainable Development Goals (Wahrhaftigkeit)

Bei der nächsten Variable *Bezugnahme zu den SDGs* handelt es sich um die einzige Variable der Subdimension SDGs. Erkennbar ist, dass der aggregierte Anteil von 32 % auf 49 % gestiegen ist. Damit referenzieren nach der CSR-RL deutlich mehr Unternehmen die SDGs. Unabhängig vom Geschäftsjahr verweist kein Medienunternehmen auf die SDGs. In der Branche Chemie & Pharmazie ergeben sich keine Änderungen und der Anteil verweilt auf einem Niveau von 50 %. Die Branche Energie & Rohstoffe steigert sich von 0 % auf 67 % erheblich, während die Branchen Maschinenbau, Bauwesen, Verkehr & Logistik moderate Zuwächse zwischen 14 % und 25 % aufweisen.

Einfachheit in der Kommunikation (Angemessenheit & Verständlichkeit)

Die Dimension Verständlichkeit & Angemessenheit besteht aus einer Subdimension: der Einfachheit in der Kommunikation. Die Subdimension beinhaltet die Variable *Lesbarkeit* sowie die *Verwendung visueller Darstellungen*.

Für die Bestimmung der Variable Lesbarkeit wurde der FRE-Score ermittelt. Für das Geschäftsjahr vor der CSR-RL beträgt das globale Minimum null und das globale Maximum 50. Die Berechnungsformel für die Normalisierung lautet demnach wie folgt: $FRE - Score(normalisiert) = \frac{FRE-Score_i}{50}$.[20] Insgesamt ist der FRE-Score von durchschnittlich 0,56 auf 0,58 um ca. 4 % gestiegen. Die Branchen Maschinenbau, Bauwesen, Verkehr & Logistik (-3 %) sowie die Branche Chemie & Pharmazie (-10 %) verzeichnen beide einen leichten Rückgang hinsichtlich der Lesbarkeit. Finanz- sowie Medienunternehmen verweilen auf einem fast konstanten Niveau, während die Branchen Handel & Konsum um 27 % und Energie & Rohstoffe um 15 % steigen.

Die Normalisierung für die Variable *Verwendung visueller Darstellungen* entfällt. Da die Variable mittels einer Kennzahl (Anzahl visueller Seiten dividiert durch betrachtete Seitenanzahl) berechnet wird, die zwischen null und eins liegt und zeitgleich das globale Minimum sowie das globale Maximum zwischen null und eins liegt. Damit können die angegebenen Werte wie folgt interpretiert werden: Wenn bspw. der Wert der Kennzahl 0,5 ist, bedeutet das, dass der Anteil der visuellen Seiten 50 % beträgt. Im Glaubwürdigkeitsindex wird analog eine Bewertung von 0,5 Punkten vorgenommen. Der Anteil visueller Darstellungen steigt nach Einführung der CSR-RL marginal um 3 %-Punkte von 50 % auf 53 %. Die Verwendung von visuellen Darstellungen sinkt in der Medienbranche von 63 % auf 33 % und halbiert sich damit fast. Auf der anderen Seite steigern bzw. verschlechtern sich die Branchen Handel & Konsum, Chemie & Pharmazie sowie Energie & Rohstoffe kaum. Finanzunternehmen (+12 %) als auch die Branche Maschinenbau, Bauwesen, Verkehr & Logistik (+15 %) verbessern sich hingegen moderat.

Insgesamt bleibt jedoch festzuhalten, dass die Streuung der Durchschnittswerte der Lesbarkeit gering ausfällt. Bei Be-

[20] Hinweis: Mit Hilfe dieser Formel kann ausgehend von dem normalisierten FRE-Score aus Tabelle 6 der durchschnittliche FRE-Score berechnet werden: FRE-Score (normalisiert) x 50 = Durchschnittlicher FRE-Score (gesamt bzw. nach Branche).

trachtung der Variable Lesbarkeit (nach Branche) liegen alle Werte zwischen 0,51 und 0,66. Damit unterscheiden sich die Branchen kaum (und entsprechend die Unternehmen) kaum. Andererseits liegt eine größere Streuung bei der Verwendung visueller Darstellungen vor. Der minimale Durchschnittswert beträgt 26 % und der maximale Durchschnittswert 70 %.

Die Ergebnisse der Bewertung jedes einzelnen Unternehmens in den Dimensionen Wahrheit, Wahrhaftigkeit sowie Angemessenheit & Verständlichkeit sind in Tabelle 7 dargestellt. Dabei werden die Bewertungen vor und nach der Einführung der CSR-RL unmittelbar gegenübergestellt. Die Summation aller drei Dimensionen entspricht schlussendlich der finalen Glaubwürdigkeitsbewertung. In der Dimension Wahrheit haben sich fünf Unternehmen (12 %) verschlechtert, ein Unternehmen ist konstant geblieben (2 %) und 35 Unternehmen (86 %) konnten sich verbessern. In Bezug auf die Wahrhaftigkeit haben sich sechs Unternehmen (15 %) verschlechtert, acht Unternehmen (20 %) haben sich weder verschlechtert noch verbessert und die restlichen 27 Unternehmen haben sich verbessert (65 %). 24 der Unternehmen (59 %) sind im Kontext der Dimension Angemessenheit & Verständlichkeit besser geworden. Die anderen 17 Unternehmen (41 %) kommen auf eine geringere Bewertung. Bei Betrachtung des gesamten Glaubwürdigkeitsindex sind fünf (12 %) Unternehmen unglaubwürdiger in der NBE geworden und die übrigen 36 (88 %) Unternehmen sind glaubwürdiger geworden.

In Abbildung 9 ist ein gestapeltes Säulendiagramm zu sehen. Dargestellt werden die jeweiligen Mittelwerte nach Branche, die die Ergebnisse aus Tabelle 7 in komprimierter Form visualisieren.

Grundsätzlich wird deutlich, dass das Geschäftsjahr, in dem die CSR-RL gültig ist (nachher), stets eine glaubwürdigere NBE besitzt als das vorausgegangene Geschäftsjahr (vorher). Die Chemie- und Pharmaziebranche verbessert sich nur um 0,66 Punkte nach Einführung der CSR-RL und verweilt folglich auf ähnlichem Niveau zum Vorjahr. Alle anderen Branchen verzeichnen deutliche Zuwächse von über 3 bis zu 6,46 Punkten nach Einführung der CSR-RL. Hervorzuheben ist insbesondere die Branche Energie & Rohstoffe, welche sich vom 5. Platz vor der CSR-RL auf den 1. Platz nach der Gesetzesreform im relativen Vergleich steigert. Andererseits entwickelt sich die Branche Chemie & Pharmazie von der Spitzenposition zu einer mittelmäßig glaubwürdigen NBE. Auffällig neben der Glaubwürdigkeitssteigerung in Abhängigkeit von der Zeit, ist, dass die umweltsensitiven Branchen, wie z.B. Energie & Rohstoffe, Maschinenbau, Bauwesen, Verkehr & Logistik sowie Chemie & Pharmazie, eine höhere Glaubwürdigkeit aufweisen als die Branchen mit einer niedrigen (bis mittleren) Umweltsensitivität. Die gilt für beide Geschäftsjahre mit der Ausnahme, dass vor der CSR-RL die Branche Energie & Rohstoffe eine relativ geringe Glaubwürdigkeit in der NBE aufweist. Neben der Glaubwürdigkeit im Gesamten sind in Abbildung 9 ebenfalls die einzelnen Dimensionen ersichtlich. Alle Branchen bis auf die Chemie- und Pharmaziebranche besitzen in der Dimension Wahrheit Zuwächse in Höhe von 1,28 bis 3,33 Punkten. In der Branche Chemie & Pharmazie kommt es zu einem leichten Minus von 0,09 Punkten. Die Dimension Wahrhaftigkeit steigt in allen Branchen mit zwischen 2 bis 3 Punkten nach dem Inkrafttreten der CSR-RL. In der Dimension Angemessenheit & Verständlichkeit ist kein eindeutiger Trend zu erkennen. Während die Medienbranche (-0,27 Punkte) sowie die Chemie- & Pharmaziebranche (-0,09) in der Bewertung fallen, zeigen die anderen Branchen ein Plus von 0,1 bis 0,15 Punkten.

In Tabelle 8 sind verschiedene Lage- und Streuparameter dargestellt, die einen deskriptiven ex-ante-ex-post-Vergleich hinsichtlich der Glaubwürdigkeit zulassen. Die Zahlen beziehen sich auf alle enthaltenen Unternehmen (unabhängig von der Branche) in der Stichprobe.

Der maximale Wert des Glaubwürdigkeitsindex beträgt 16,65 Punkte und das Minimum beträgt null Punkte. Durchschnittlich steigt die Glaubwürdigkeit der NBE um 3,46 Punkte. Der Anstieg besteht aus 1,55 Punkte aus der Dimension Wahrheit, 1,85 Punkten aus der Dimension Wahrhaftigkeit sowie 0,06 Punkten aus der Dimension Angemessenheit & Verständlichkeit. Daran kann man erkennen, dass die gestiegene Glaubwürdigkeit vorrangig aus den Dimensionen Wahrheit und Wahrhaftigkeit resultiert. Dies ist allerdings aufgrund der geringen Gewichtung der Dimension Angemessenheit & Verständlichkeit wenig verwunderlich. Ferner ist anzumerken, dass die Standardabweichung nach der Einführung der CSR-RL in allen Dimensionen sinkt. Somit fällt die durchschnittliche Abweichung vom Durchschnitt (Streuung) geringer aus. Dies spricht für eine Angleichung der Glaubwürdigkeit zwischen den Unternehmen. Mit anderen Worten ausgedrückt: Es gibt nach der CSR-RL weniger High- und weniger Low-Performer und damit gibt es weniger Streuung um den Mittelwert.

In Abbildung 10 ist eine Wahrheit-Wahrhaftigkeits-Matrix dargestellt. Das Ausklammern der Dimension Angemessenheit & Verständlichkeit erscheint aus dem Grund sinnvoll, da die Gewichtung im Glaubwürdigkeitsindex mit ca. 10 % gering ausfällt und kein klarer Trend in dieser Dimension zu erkennen ist (s. z.B. Abb. 8). Auf der Abszisse ist die Wahrheit abgetragen und auf der Ordinate befindet sich die Wahrhaftigkeit. Für beide betrachtete Geschäftsjahre ist eine Mittelwertlinie eingezeichnet. Die blauen Mittelwertlinien (Wahrheit = 2,46; Wahrhaftigkeit = 3,95; N = 41) beziehen sich auf die Durchschnittswerte vor der Einführung der CSR-RL, wohingegen die orangen Mittelwertlinien (Wahrheit = 4,01; Wahrhaftigkeit = 5,80; N = 41) sich auf die Geschäftsjahre danach beziehen. Bei Betrachtung der blauen Mittelwertlinien liegen vor der CSR-RL ca. 46 % der Unternehmen im ersten Quadranten. Nach der CSR-RL steigt der Anteil um 30 %-Punkte auf 76 % der Unternehmen. Wenn die orangenen Mittelwertlinien hinzugezogen werden, liegen noch 49 % nach der CSR-RL über den Durchschnittswerten und somit im ersten Quadranten. Vor der CSR-RL sind es hingegen lediglich ca. 20 %, die im ersten Quadranten liegen. Demnach zeichnet sich in der Abbildung 10 eine deutliche Verbesserung in den Dimensionen Wahrheit und Wahrhaftigkeit ab. Außerdem wird deutlich, dass es zehn Unternehmen gibt, die in beiden Dimensionen mit null Punkten (bzw. nahe der

Tabelle 7: Glaubwürdigkeitsindex im ex-ante-ex-post-Vergleich auf Unternehmensebene, N = 41 (eigene Darstellung)

Unternehmen	CSR-RL	W	WH	AV	GW
MASCHINENBAU, BAUWESEN, VERKEHR & LOGISTIK (N = 13)					
Brenntag AG	Vorher	3,26	6,00	1,28	10,54
	Nachher	4,35	6,00	1,52	11,87
Fraport AG	Vorher	4,20	6,00	1,74	11,94
	Nachher	6,67	6,00	1,54	14,21
GEA Group AG	Vorher	2,11	4,00	0,90	7,01
	Nachher	3,16	4,00	1,14	8,30
Hella GmbH & Co. KGaA	Vorher	0,01	0,00	0,91	0,93
	Nachher	3,04	2,00	0,55	5,58
HOCHTIEF AG	Vorher	3,95	8,00	1,48	13,43
	Nachher	4,95	9,00	1,42	15,37
Jungheinrich AG	Vorher	3,86	7,00	1,54	12,40
	Nachher	4,86	6,00	1,58	12,44
KION Group AG	Vorher	3,46	4,00	1,08	8,54
	Nachher	5,74	7,00	0,94	13,68
LEONI AG	Vorher	1,80	1,00	0,72	3,52
	Nachher	2,17	8,00	1,06	11,23
MTU Aero Engines AG	Vorher	4,15	7,00	1,20	12,35
	Nachher	4,10	8,00	1,56	13,66
NORMA Group SE	Vorher	3,67	8,00	1,42	13,09
	Nachher	4,99	8,00	1,68	14,67
Osram Licht AG	Vorher	4,18	5,00	0,86	10,04
	Nachher	4,91	7,00	1,22	13,13
Rheinmetall AG	Vorher	0,01	0,00	0,40	0,41
	Nachher	4,43	4,00	1,36	9,79
Schaeffler AG	Vorher	4,10	8,00	1,38	13,48
	Nachher	4,71	7,00	1,04	12,75
Mittelwert	Vorher	2,98	4,92	1,15	9,05
	Nachher	4,47	6,31	1,28	12,05
FINANZEN (N = 8)					
Aareal Bank AG	Vorher	4,15	8,00	0,92	13,07
	Nachher	3,10	7,00	0,90	11,00
Deutsche Wohnen SE	Vorher	4,39	6,00	1,16	11,55
	Nachher	5,44	5,00	1,34	11,78
Hannover Rück SE	Vorher	3,92	5,00	1,08	10,00
	Nachher	4,27	7,00	0,92	12,19
LEG Immobilien AG	Vorher	0,02	0,00	0,58	0,60
	Nachher	5,59	7,00	1,14	13,73
Deutsche Pfandbriefbank AG	Vorher	0,04	1,00	0,56	1,60
	Nachher	4,13	4,00	0,60	8,73
Rocket Internet SE	Vorher	0,01	0,00	0,00	0,01
	Nachher	3,10	1,00	0,34	4,44
TAG Immobilien AG	Vorher	0,08	0,00	0,88	0,96
	Nachher	4,09	6,00	0,92	11,01
Talanx AG	Vorher	4,03	4,00	1,02	9,05
	Nachher	3,76	6,00	0,88	10,64
Mittelwert	Vorher	2,08	3,00	0,78	5,86
	Nachher	4,18	5,38	0,88	10,44

(Continued)

Table 7—continued

HANDEL & KONSUM (N = 7)					
Ceconomy AG	Vorher	0,06	2,00	1,34	3,40
	Nachher	2,22	6,00	0,80	9,02
Delivery Hero AG	Vorher	0,01	0,00	1,02	1,03
	Nachher	0,04	3,00	1,37	4,41
Fielmann AG	Vorher	0,00	0,00	0,00	0,00
	Nachher	3,26	4,00	1,16	8,42
HUGO BOSS AG	Vorher	4,36	9,00	1,26	14,62
	Nachher	4,52	9,00	1,20	14,72
METRO AG	Vorher	3,50	2,00	1,26	6,76
	Nachher	5,84	8,00	1,60	15,44
PUMA SE	Vorher	3,15	5,00	1,66	9,81
	Nachher	3,17	5,00	1,70	9,87
Zalando SE	Vorher	2,50	5,00	1,20	8,70
	Nachher	3,48	5,00	1,00	9,48
Mittelwert	Vorher	1,94	3,29	1,11	6,33
	Nachher	3,22	5,71	1,26	10,19
CHEMIE & PHARMAZIE (N = 6)					
Evonik Industries AG	Vorher	5,58	7,00	1,26	13,84
	Nachher	5,60	9,00	1,40	16,00
FUCHS PETROLUB SE	Vorher	1,04	0,00	1,29	2,33
	Nachher	0,06	3,00	1,16	4,22
Gerresheimer AG	Vorher	0,06	1,00	0,90	1,96
	Nachher	0,08	3,00	0,96	4,04
K+S AG	Vorher	2,40	8,00	0,86	11,26
	Nachher	2,40	7,00	0,96	10,36
LANXESS AG	Vorher	3,87	9,00	1,46	14,33
	Nachher	4,23	8,00	1,10	13,33
Symrise AG	Vorher	7,57	8,00	1,08	16,65
	Nachher	7,63	8,00	0,74	16,37
Mittelwert	Vorher	3,42	5,50	1,14	10,06
	Nachher	3,33	6,33	1,05	10,72
MEDIEN (N = 4)					
CTS Eventim KG & Co. KGaA	Vorher	0,00	0,00	0,00	0,00
	Nachher	3,15	4,00	0,56	7,71
ProSiebenSat.1 Media SE	Vorher	3,50	4,00	1,88	9,38
	Nachher	5,99	5,00	1,36	12,35
Scout24 AG	Vorher	2,60	5,00	1,06	8,66
	Nachher	5,00	7,00	1,24	13,24
Ströer SE & Co. KGaA	Vorher	0,01	0,00	2,00	2,01
	Nachher	1,09	1,00	0,72	2,81
Mittelwert	Vorher	1,53	3,00	0,90	5,01
	Nachher	3,81	6,00	1,03	9,03
ENERGIE & ROHSTOFFE (N = 3)					
Aurubis AG	Vorher	0,39	2,00	1,02	3,41
	Nachher	4,61	6,00	1,34	11,95
innogy SE	Vorher	4,83	7,00	0,68	12,51
	Nachher	4,78	8,00	1,00	13,78
Salzgitter AG	Vorher	0,02	0,00	1,00	1,02
	Nachher	5,85	4,00	0,76	10,61
Mittelwert	Vorher	1,75	3,00	0,90	5,65
	Nachher	5,08	6,00	1,03	12,11

W = Wahrheit, WH = Wahrhaftigkeit, AV = Angemessenheit & Verständlichkeit, GW = Glaubwürdigkeitsindex

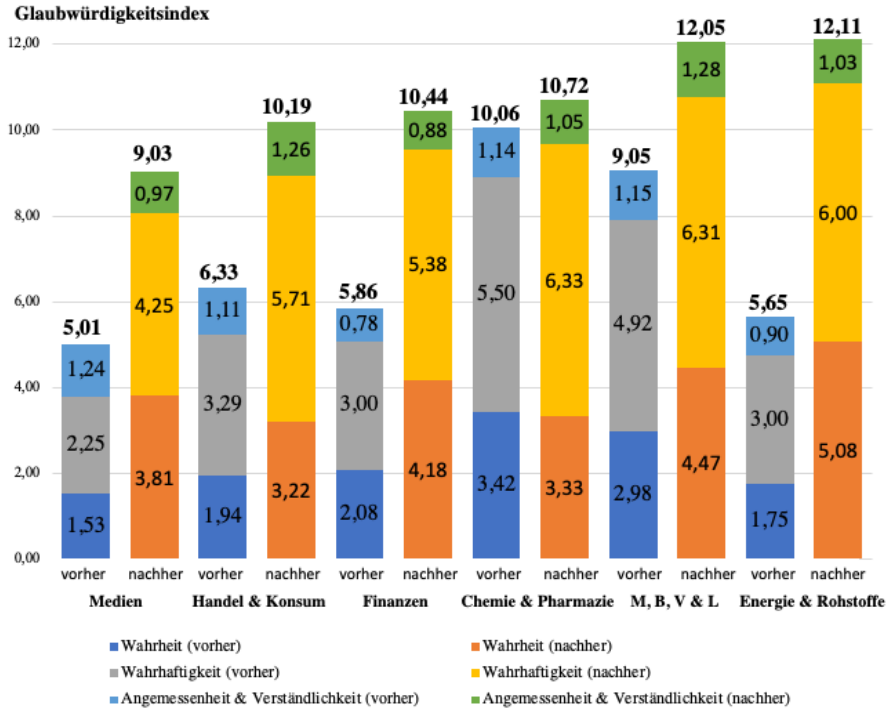


Abbildung 9: Mittelwertvergleich nach Branche, N=41 (eigene Darstellung)

Tabelle 8: Lage- und Streumaße zum Glaubwürdigkeitsindex im ex-ante-ex-post-Vergleich, N = 41 (eigene Darstellung)

Dimension	CSR-RL	Min.	Max.	Mittelwert	Standardabweichung
Wahrheit	Vorher	0,00	7,57	2,46	2,02
	Nachher	0,04	7,63	4,01	1,71
Wahrhaftigkeit	Vorher	0,00	9,00	3,95	3,21
	Nachher	1,00	9,00	5,80	2,13
Angemessenheit & Verständlichkeit	Vorher	0,00	2,00	1,06	0,46
	Nachher	0,34	1,70	1,12	0,33
Glaubwürdigkeitsindex	Vorher	0,00	16,65	7,47	5,30
	Nachher	2,81	16,37	10,93	3,54

null Punkte) unterdurchschnittlich schlecht abschneiden.

6.5.2. Induktive Ergebnisse

Nachdem im vorangegangenen Kapitel bereits erste Aussagen deskriptiver Natur vorgenommen werden konnten, die für eine Glaubwürdigkeitssteigerung nach der Einführung der CSR-RL sprechen, wird im Folgenden mittels induktiver Statistik die Signifikanz der Ergebnisse überprüft. Zunächst erfolgt eine nichtparametrische Signifikanzanalyse auf der Variablenebenen und der Glaubwürdigkeit als Ganzes (inklusive der einzelnen Dimensionen). Danach wird als parametrisches Verfahren eine multiple lineare Regressionsanalyse verwendet, um für mögliche Störgrößen zu kontrollieren.

Nichtparametrische Verfahren

In den deskriptiven Ergebnissen ist explizit auf die Branchenunterschiede eingegangen worden. Die branchenspezifische Betrachtungsweise mit nicht-parametrischen Verfahren scheidet aufgrund zu geringer Beobachtungszahlen (N < 20) aus. Des Weiteren zeigen die deskriptiven Ergebnisse, dass es einige Ausreißer gibt, die die Ergebnisse möglicherweise verzerren. Aus diesem Grund wird für die Robustheit der Ergebnisse zusätzlich eine Analyse ohne Ausreißer erfolgen.

Nachfolgend werden zunächst die Variablen auf signifikante Unterschiede vor und nach der CSR-RL untersucht. Für die Testung wird entweder der Wilcoxon-Vorzeichen-Rang-Test[21] für metrische bzw. ordinalskalierte Variablen oder ein

[21] Da die Voraussetzung für einen t-Test nicht gegeben ist (Normalverteilungsannahme der Stichprobendifferenzen ist verletzt), wird der nichtparametrische Wilcoxon-Vorzeichen-Rang-Test verwendet. Dieser testet, ob sich

Abbildung 10: Wahrheit-Wahrhaftigkeitsmatrix auf der Unternehmensebene (in Anlehnung an Mazzotta et al., 2020, S. 1910)

Tabelle 9: Nichtparametrische Signifikanzanalyse auf Variablenebene, N = 41 (eigene Darstellung)

Variable	Mittelwert vorher	Mittelwert nachher	Differenz	p-Wert	Teststatistik
Anbieter	0,29	0,73	0,44	0,000***	Z = 171,00
Prüfungsumfang	0,05	0,32	0,27	0,001***	$\chi^2 = 9,09$
Prüfungssicherheit	0,00	0,00	0,00	-	-
Verwendung eines Rahmenwerk	0,61	0,80	0,19	0,011***	Z = 49,50
Länge des Nachhaltigkeitsberichts (normalisiert)	0,21	0,30	0,09	0,000***	Z = 668,00
Verortung der NFI	0,49	0,73	0,24	0,006***	$\chi^2 = 6,75$
Methodik-Kapitel	0,56	0,78	0,22	0,012***	$\chi^2 = 5,81$
Genauigkeit der dargestellten Informationen	0,25	0,35	0,10	0,034**	Z = 96,00
Existenz einer Wesentlichkeitsmatrix	0,37	0,44	0,07	0,453	$\chi^2 = 5,71$
Aktualität der Wesentlichkeitsanalyse	0,39	0,83	0,44	0,000***	$\chi^2 = 13,14$
Bekenntnis des Vorstands zur Nachhaltigkeit	0,59	0,80	0,21	0,012***	$\chi^2 = 5,82$
Identifikation	0,59	0,88	0,29	0,000***	$\chi^2 = 10,08$
Dialog	0,61	0,85	0,24	0,006***	$\chi^2 = 6,75$
Engagement	0,17	0,17	0,00	1,00	$\chi^2 = 0,00$
Organisationale Verankerung	0,93	1,34	0,41	0,004***	Z = 45,00
Bezugnahme zu den SDGs	0,32	0,49	0,17	0,039**	$\chi^2 = 4,00$
Lesbarkeit	0,56	0,58	0,02	0,984	Z = 411,50
Verwendung visueller Darstellungen	0,50	0,53	0,03	0,401	Z = 366,00

Statistische Signifikanz auf dem 1% (5%, 10%) Level wird bezeichnet durch *** (**,*).

der Median der Differenzen der betrachteten Variablen von null unterscheidet. Die Nullhypothese ist, dass dem nicht so ist. Grundsätzlich untersucht der Wilcoxon-Vorzeichen-Rang-Test also, ob sich die zentralen Tendenzen zwischen den zwei Zeitpunkten unterscheiden.

McNemar-Test[22] für binäre Variablen verwendet. Beide Tests sind für die Überprüfung von abhängigen bzw. verbundenen Stichproben entwickelt worden. In Tabelle 9 sind die Ergebnisse illustriert. Da beide verwendete Tests nur aussagen, dass sich beide Stichproben signifikant voneinander unterscheiden, ist zusätzlich noch der Mittelwert angegeben, um den Richtungseffekt einschätzen zu können.[23]

Der McNemar-Test kann nicht für die Variable *Prüfungssicherheit* durchgeführt werden, da kein Unternehmen eine externe Prüfung mit hoher Prüfungssicherheit durchführen lässt. Alle Variablen weisen eine positive Differenz auf. Das heißt, dass sich alle Variablen im Durchschnitt verbessert haben. Die Variable *Genauigkeit der dargestellten Informationen* ist auf dem 5 %-Signifikanzniveau, während alle anderen Variablen (außer die Variablen: *Existenz einer Wesentlichkeitsmatrix*, *Engagement*) hochsignifikant auf dem 1 % Level sind. Die Variable *Existenz einer Wesentlichkeitsmatrix* unterscheidet sich in den beiden Jahren nicht signifikant voneinander ($\chi^2 = 5,71, p = 0,453$). Außerdem bleibt die Häufigkeitsverteilung der Variable *Engagement* 1:1 vor und nach der CSR-RL identisch. Entsprechend liegt keine statistische Signifikanz vor ($\chi^2 = 0,00, p = 1,00$).

Wie in der deskriptiven Statistik (s. Kap. 6.5.1) zum Beispiel in Tabelle 7 oder Abbildung 9 deutlich wurde, gibt es einige Unternehmen, die nach der Einführung der CSR-RL erstmalig eine NBE besitzen. Um diesen Effekt durch die First-time-Reporter (FTR) auszuklammern, wird eine nichtparametrische Signifikanzanalyse ohne die FTR vorgenommen. FTR sind wie folgt definiert: Ein FTR hat vor dem Inkrafttreten der CSR-RL weniger als fünf Seiten NFI publiziert.[24] Insgesamt reduziert sich damit die Stichprobe auf 33 Unternehmen. Im Anhang in Tabelle 13 sind die Ergebnisse abgebildet. Die Variable *Verwendung eines Rahmenwerks* ist nicht mehr hoch, sondern nur noch schwach signifikant. Zusätzlich sind die Variablen *Verortung der NFI*, *Methodik-Kapitel*, *Bekenntnis des Vorstands zur Nachhaltigkeit* sowie *Dialog* gar nicht mehr signifikant, obwohl alle zuvor hochsignifikant waren. Durch die Exklusion der FTR relativieren sich die Ergebnisse etwas. Bei allen Unternehmen unterscheiden sich elf (der insgesamt 18 Variablen) hochsignifikant und zwei signifikant voneinander. Nach der Exklusion der FTR sind es nur noch sechs hochsignifikante, ein signifikanter und zwei schwach signifikante Unterschiede feststellbar. Trotzdem bleibt allgemein festzuhalten, dass die nichtparametrische Testung die Hypothese, dass die CSR-RL die Glaubwürdigkeit in den Dimensionen Wahrheit und Wahrhaftigkeit verbessert, untermauert. Die Ergebnisse aus der deskriptiven Statistik haben sich mittels induktiver Statistik nochmals verstärkt. Das neue Gesetz hat nachweislich keinen Einfluss auf die Dimension Angemessenheit & Verständlichkeit.

In Tabelle 10 ist nochmal der gesamte Glaubwürdigkeitsindex samt seiner drei Dimensionen dargestellt. Da sowohl der Glaubwürdigkeitsindex als auch die einzelnen Dimensionen metrisch skaliert sind, wird der Wilcoxon-Vorzeichen-Rang-Test angewendet. Erneut wird die Stichprobe einmal komplett betrachtet (N = 41) und einmal ohne die FTR (N = 33).

In allen Fällen ist ein positiver Richtungseffekt zu erkennen. Demnach handelt es sich stets um eine durchschnittliche Verbesserung in den einzelnen Dimensionen und somit auch insgesamt in der Glaubwürdigkeit. Beim Vergleich der Mittelwerte vor der CSR-RL in beiden Stichproben, ist zu erkennen, dass die FTR als Ausreißer erwartungsgemäß den Durchschnitt nach unten verzerren. Entsprechend sind die Durchschnittswerte mit FTR niedriger als die Mittelwerte ohne FTR. Sowohl die Dimension Wahrheit als auch die Dimension Wahrhaftigkeit sind hochsignifikant verschieden voneinander unabhängig davon, ob die FTR inkludiert sind oder nicht. Bei der Dimension Angemessenheit & Verständlichkeit liegt keine statistisch signifikante Verschiedenheit vor und nach Einführung der CSR-RL sowohl mit als auch ohne FTR vor. Insgesamt steigt der Glaubwürdigkeitsindex auf dem 1 %-Signifikanzniveau mit als auch ohne Ausreißer.

Parametrische Verfahren

Um herauszufinden, ob möglicherweise andere Faktoren ursächlich für die Glaubwürdigkeitssteigerung sind, wird zusätzlich ein parametrisches Verfahren verwendet. Da es sich um Daten mit 41 Querschnittseinheiten (N = 41) und zwei Zeitreihenlängeneinheiten (T = 2), kommen drei verschiedene Modelle für eine multiple lineare Regression in Frage: (a) ein Pooled OLS Modell (POLS), (b) ein Fixed Effects Modell (FEM) sowie (c) ein Random Effects Modell (REM). Mittels der Modellpaneldiagnostik[25] ergibt sich für den Glaubwürdigkeitsindex sowie für die einzelnen Dimensionen, dass ein REM das geeignetste Modell darstellt. In Tabelle 11 sind die resultierenden Regressionsmodelle dargestellt.

Neben den bereits vorgestellten Kontrollvariablen Unternehmensgröße (SIZE), Ertragslage (ROA), Verschuldungsgrad (LEV), Eigentumsstreuung (DISP), Medienpräsenz (PRES) sowie der Umweltsensitivität (ENVSEN) wird eine weitere Zeitdummy-Variable eingeführt. Diese ist gleich

[22] Der McNemar-Test gehört zu der Chi-Quadrat-Test-Gruppe und wird für die Prüfung statistischer Signifikanz bei dichotomen Variablen verwendet. Die Nullhypothese besagt, dass sich die beiden Stichproben nicht in der Häufigkeitsverteilung unterscheiden.

[23] Hinweis: Die Multiplikation des Mittelwerts mit 41 (N = 41) ergibt die kumuliert erreichte Punktzahl in der jeweiligen Variable.

[24] Die genaue Abgrenzung von gewerteten Seiten, die den NFI angehören, befindet sich in der Excel-Datei „Datenerhebung“ unter dem Reiter „NFI im Sample“. Zu den FTR gehören folgende Unternehmen: CTS Eventim AG & Co. KGaA, Delivery Hero AG, Fielmann AG, Hella GmbH & Co. KGaA, LEG Immobilien AG, Rheinmetall AG, Rocket Internet SE sowie die Ströer SE & Co. KGaA.

[25] In Gretl werden folgende Tests zur Ermittlung des optimalen Regressionsmodells durchgeführt: (a) POLS vs. FEM - F-Test (gemeinsame Signifikanz der verschiedenen Gruppenmittel), (b) POLS vs. REM - Breusch-Pagan-Test sowie (c) REM vs. FEM - Hausman-Test.

[26] Es wird ein linearer Zusammenhang zwischen den erklärenden und den zu erklärenden Variablen aufgrund der bestehenden Literatur angenommen. Die Gauß-Markow-Annahmen wurden geprüft. In der Dimension Angemessenheit & Verständlichkeit liegt Heteroskedastizität vor. Ferner ist die Voraussetzung, dass eine Normalverteilung der Residuen gegeben ist, in der Dimension Wahrhaftigkeit verletzt. Demnach sind die Ergebnisse möglicherweise verzerrt.

Tabelle 10: Nichtparametrische Signifikanzanalyse der Glaubwürdigkeit mit und ohne FTR (eigene Darstellung)

	Glaubwürdigkeitsindex Mit FTR (N = 41)				Glaubwürdigkeitsindex Ohne FTR (N = 33)			
	W	WH	AV	GW	W	WH	AV	GW
Mittelwert (vorher)	2,46	3,95	1,06	7,47	3,05	4,91	1,16	9,12
Mittelwert (nachher)	4,01	5,80	1,12	10,93	4,27	6,42	1,17	11,86
Differenz	1,55	1,85	0,06	3,46	1,22	1,15	0,01	2,74
p-Wert	0,00***	0,00***	0,297	0,00***	0,00***	0,00***	0,795	0,00***
Teststat. Z	799	522	511	811	504	292	295	514

Statistische Signifikanz auf dem 1% (5%, 10%) Level wird bezeichnet durch *** (**,*).

Tabelle 11: Random Effects Modelle, N = 82 (eigene Darstellung) [26]

Robuste (HAC) Standardfehler in Klammern. Es handelt sich um ein REM mit GLS-Schätzern. Die angegebenen Zahlen sind auf drei Nachkommastellen gerundet. Die Begründung für die Aufnahme und die Messung der Kontrollvariablen sind im Kapitel 6.3 erklärt. Zusätzlich wird die Variable Zeitdummy eingeführt. Diese ist gleich 1, wenn das Geschäftsjahr nach der CSR-RL ist. Statistische Signifikanz auf dem 1% (5%, 10%) Level wird bezeichnet durch *** (**,*).

	Wahrheit	Wahrhaftigkeit	Angemessenheit & Verständlichkeit	Glaubwürdigkeitsindex
Zeitdummy	1,484***	1,778***	0,053	3,313***
	(0,280)	(0,343)	(0,067)	(0,580)
SIZE	1,033***	0,922**	0,052	2,022***
	(0,178)	(0,335)	(0,056)	(0,502)
ROA	0,055*	-0,001	0,007	0,062
	(0,029)	(0,035)	(0,011)	(0,056)
LEV	-0,141***	0,001*	-0,017	-0,162
	(0,045)	(0,112)	(0,014)	(0,151)
DISP	0,841**	0,635***	0,097	2,980***
	(0,401)	(0,607)	(0,101)	(0,969)
PRES	0,001	0,004	0,000	0,005
	(0,001)	(0,002)	(0,000)	(0,004)
ENVSEN	0,637	2,056***	0,157	2,833***
	(0,407)	(0,687)	(0,128)	(1,083)
Konstante	-7,373***	-6,794***	0,463	-13,800***
	(1,379)	(2,578)	(0,554)	(3,853)
Beobachtungen	82	82	82	82

eins, wenn das Geschäftsjahr nach der CSR-RL betrachtet wird. Im Folgenden wird hauptsächlich auf den Glaubwürdigkeitsindex im Gesamten eingegangen. Die anderen drei Regressionsmodelle, die jeweils die Dimension als abhängige Variable haben, dienen dazu, zu eruieren, worin die Glaubwürdigkeitsveränderung herrührt. Festzuhalten ist vorab, dass analog zu den bisherigen Ergebnissen alle potentiellen Einflussgrößen auf die Dimension Angemessenheit & Verständlichkeit keinen statistischen signifikanten Einfluss besitzen. Damit ergeben sich alle statistische Signifikanzen aus den Dimensionen Wahrheit und Wahrhaftigkeit.

Die Glaubwürdigkeit der NBE ist nach der CSR-RL durchschnittlich um 3,313 Punkte höher verglichen mit der Glaubwürdigkeit vor der Einführung der CSR-RL ceteris paribus (c.p.). Dies liefert zusätzliche Evidenz für die aufgestellte Hypothese, dass die Glaubwürdigkeit nach der Einführung der CSR-RL steigen wird. Anhand der Regressionsmodelle wird jedoch ebenso deutlich, dass die kontrollierten Einflussgrößen auch teilweise hochsignifikante Auswirkungen auf die Glaubwürdigkeit besitzen. Wenn die Unternehmensgröße um 1 % steigt, erhöht sich die Glaubwürdigkeit der NBE im Durchschnitt um 0,02 Einheiten (c.p.). Ferner haben Unternehmen mit einer hohen Eigentumsstreuung eine durchschnittlich um 2,980 Punkte glaubwürdigere NBE als Unternehmen mit einer geringen Eigentumsstreuung (c.p.). Zudem zeigt sich, dass Unternehmen mit einer hohen Umweltsensitivität eine um 2,833 Punkte durchschnittlich glaubwürdigere NBE haben als Unternehmen mit einer geringen bzw. mittleren Umweltsensitivität (c.p.). Damit lässt sich der Einfluss der Umweltsensitivität, der sich bereits in der deskriptiven statistischen Auswertung zeigt, induktiv verifizieren. An dieser Stelle ist noch hervorzuheben, dass in der Dimension

Wahrheit die erwarteten Effekte hinsichtlich der Ertragslage und der Verschuldung eingetreten sind. Eine profitable Lage führt folglich zu einer glaubwürdigeren NBE, während eine hohe Verschuldung die Glaubwürdigkeit in der NBE schmälert. Die Kontrollvariable Medienpräsenz zeigt keine bedeutsame Auswirkung auf die Glaubwürdigkeit der NBE. In Tabelle 14 im Anhang ist die gleiche Regressionsanalyse ohne FTR durchgeführt worden. Alle genannten Effekte bewahren ihre statistische Signifikanz, wenngleich die Umweltsensitivität nur noch schwach signifikant ist. Damit sind die Ergebnisse als robust einzuschätzen. Um die Ergebnisse weiter zu verifizieren, wurden zwei Differenzregressionsanalysen mit und ohne FTR durchgeführt (s. Tab. 15 und Tab. 16 im Anhang). Für alle Variablen wurden Differenzvariablen gebildet, in dem das Geschäftsjahr nach der CSR-RL mit dem Geschäftsjahr danach verrechnet wird.[27] Die (weitgehend) fehlende statistische Signifikanz zeigt auf, dass die Veränderung des Glaubwürdigkeitsniveaus in der NBE nicht aus der Veränderung potentieller Störeinflüsse entsteht.

6.5.3. Diskussion der Ergebnisse

Im Folgenden werden insbesondere die wesentlichen Ergebnisse vor dem Hintergrund der CSR-RL diskutiert. Aufgrund der unterschiedlichen Stichprobe und der teils anderen Herangehensweise in der Inhaltsanalyse ist ein Vergleich zu Mazzotta et al. (2020) nur an bestimmten Stellen sinnvoll.

Überraschend ist festzuhalten, dass deutlich mehr Unternehmen eine externe Prüfung durchgeführt haben, obgleich dies nicht explizit in der deutschen Gesetzgebung vorgeschrieben ist. Dieses Resultat kann vermutlich damit begründet werden, dass die Unternehmen aufgrund der Neuheit des Gesetzes (und der damit verbundenen Unsicherheit über die neuen Anforderungen) und den erhöhten Sanktionen bereit waren, finanzielle Mittel für eine umfassende externe Prüfung bereitzustellen (DNK, 2018, S. 23). Obgleich kein „Hard Law" besteht, also keine gesetzliche Verpflichtung zur externen Prüfung, zeigt dieses Ergebnis, dass die CSR-RL trotzdem zu einer Verbesserung geführt hat. Dieses Ergebnis liefert ein empirisches Argument für die Fürsprecher von „Soft Laws", die im Gegensatz zu „Hard Laws" nicht verpflichtend einzuhalten sind (Schwarze, 2011, S. 3). Im Kontext der externen Prüfung kann dem dennoch entgegnet werden, dass zwar 70 % Anteil an Unternehmen eine deutliche Steigerung darstellt, allerdings ein „Hard Law" wie in Italien unmittelbar zu einem noch höheren Anteil extern geprüfter Unternehmen führt (Mazzotta et al., 2020, S. 1909). Es besteht allerdings noch Verbesserungspotential hinsichtlich des Prüfungsumfangs und der Prüfungssicherheit. Gemäß der Inhaltsanalyse von Mazzotta et al. (2020, S. 1909) wurde der Inhalt der extern geprüften Inhalte nicht spezifiziert bzw. war nie vollumfänglich. Wenngleich die empirischen Ergebnisse der Masterarbeit in der Variable *Prüfungsumfang* bessere Ergebnisse aufzeigt, können die Resultate trotzdem als schwach eingeordnet werden. Der Prüfungsumfang und die Prüfungssicherheit sind entscheidend dafür, dass alle NFI mit einer hohen Wahrscheinlichkeit frei von wesentlichen Falschdarstellungen sind. Die Variable *Verwendung eines Rahmenwerks* hat erheblich an Bedeutung gewonnen. An dieser Stelle zeigt sich, dass der Rechtfertigungsgedanke des § 289d HGB, nach dem ein Unternehmen begründen muss, weswegen es kein Rahmenwerk für die NBE benutzt, Wirkung zeigt. Dennoch bleibt kritisch anzumerken, dass der Comply-and-Explain Ansatz viele Schlupflöcher für Unternehmen ermöglichst, systematisch Informationen auszulassen (La Torre, Sabelfeld, Blomkvist, Tarquinio & Dumay, 2018, S. 615). Daneben ist die *Seitenlänge der NFI* und die *Genauigkeit der dargestellten Informationen* nach Einführung der CSR-RL signifikant gestiegen. Dies steht mit hoher Wahrscheinlichkeit im kausalen Zusammenhang mit der Ausweitung der Mindestinhalte der NFE. Die gesetzliche Konkretisierung, worüber und wie ein Unternehmen berichten muss, kann somit als wirkungsvoll angesehen werden. Außerdem lässt sich an der erhöhten Informationsmenge ablesen, dass die Mindestinhalte für zahlreiche Unternehmen eine Neuigkeit darstellt, über die zusätzlich berichtet werden muss. Dies wird durch die Tatsache untermauert, dass in der betrachteten Stichprobe insgesamt acht Unternehmen erstmalig nach Einführung der CSR-RL über NFI berichten.

Während sich das von der GRI empfohlene Tool der Wesentlichkeitsmatrix zur Priorisierung relevanter Themen nicht durchgesetzt hat, verfügen (fast) alle Unternehmen (83 %) nach der CSR-RL über eine aktuelle Wesentlichkeitsanalyse. Dies verdeutlicht, dass ein Großteil der untersuchten Unternehmen verinnerlicht hat, dass die Themenfindung ein dynamischer Prozess ist, der fortlaufend kritisch reflektiert werden muss. Zudem veranschaulicht der höhere Zuspruch durch den Vorstand, sich für Nachhaltigkeitsthemen einzusetzen, dass das Thema Nachhaltigkeit eine höhere Relevanz nach der Einführung der CSR-RL eingenommen hat. Folglich ist die CSR-RL in diesem Kontext als positiv einzuschätzen. Der Aspekt der Wesentlichkeit wird durch die Wesentlichkeitsformel des § 289c Abs. 3 HGB tangiert. Nach Einführung der CSR-RL schaffen es die meisten Unternehmen, relevante Stakeholdergruppen zu identifizieren. Allerdings besteht noch großes Potential, den Dialog bzw. das Engagement mit den Stakeholdern zu steigern. Für die meisten Unternehmen findet das aktive Einbinden von Stakeholdern in Entscheidungsprozesse nicht statt. Demnach hat die CSR-RL nicht maßgeblich zu einem besseren Dialog bzw. Engagement beigetragen. § 289c HGB beinhaltet ferner, dass Unternehmen darüber berichten müssen, mit welcher Governance-Struktur sie die Nachhaltigkeitsthemen angehen. Diese Publizitätspflicht der CSR-RL spiegelt sich deutlich in den Ergebnissen wider. Zunehmend mehr Unternehmen bauen eigene organisationale Strukturen für das Thema Nachhaltigkeit auf. Demnach entfaltet die CSR-RL seine Wirkung. Im Vergleich zu den untersuchten italienischen Unternehmen in der Studie von Mazzotta et al. (2020, S. 1909), ist die organisationale Verankerung in der betrachteten Stichprobe schwächer ausgeprägt. Allerdings ist positiv anzumerken, dass in der

[27] Bsp.: Diff_SIZE = Unternehmensgröße nach der CSR-RL – Unternehmensgröße vor der CSR-RL

betrachteten Stichprobe deutlich häufiger eine Organisation gegründet wurde, die speziell Themen der Nachhaltigkeit adressiert. Die *Bezugnahme zu den SDGs* ist zwar mit 32 % auf 49 % merklich mit der CSR-RL gestiegen, jedoch im Vergleich zu den italienischen Unternehmen mit 77 % als relativ gering einzustufen (Mazzotta et al., 2020, S. 1909). Da die CSR-RL keinerlei Bezugspunkt zu den SDGs aufweist, ist diese Verbesserung als überraschendes Ergebnis einzustufen.

Die Dimension Angemessenheit & Verständlichkeit bleibt durch die CSR-RL vollkommen unberührt. Dies hat sich auch in den Ergebnissen durch ein unverändertes Niveau vor und nach der CSR-RL herauskristallisiert. Beim Vergleich der Variable *Verwendung visueller Darstellungen* mit dem empirischen Befund von Mazzotta et al. (2020, S. 1910) fällt auf, dass die Resultate auf einem ähnlichen (absoluten) Niveau liegen. Insgesamt erscheint dieses Ergebnis wenig überraschend, da die CSR-RL keinerlei Berührungspunkt zu der Art und Weise der Kommunikation aufweist.

Während in der Studie von Mion und Loza Adaui (2019, S. 18) im DAX keine FTR vor der CSR-RL vorhanden waren, so wurden im MDAX acht FTR identifiziert, die nach dem Inkrafttreten des Gesetzes über NFI berichten mussten. An dieser Stelle zeigt sich die Notwendigkeit der CSR-RL, die auch die letzten „CSR-Verweigerer" zu einer NBE zwingt. Insgesamt liefert die durchgeführte Untersuchung neben den Studien von Hąbek und Wolniak (2016) sowie Mion und Loza Adaui (2019) weitere Evidenz für die Hypothese, dass eine verpflichtende NBE zu einer glaubwürdigeren Berichtserstattung führt. Insbesondere zeigen die Ergebnisse in Kombination mit der Studie von Mion und Loza Adaui (2019), dass die CSR-RL in ihrer Ausgestaltung zu einer glaubwürdigeren NBE beigetragen hat.

Summa summarum haben die Verbesserungen in den Dimensionen Wahrheit und Wahrhaftigkeit und das konstante Level der Dimension Angemessenheit & Verständlichkeit zu einer Glaubwürdigkeitssteigerung in der NBE geführt, die durch die CSR-RL induziert wurde. Dass die Glaubwürdigkeitssteigerung in der NBE durch interne oder externe Einflüsse zurückzuführen ist, konnte durch parametrische Verfahren weitgehend ausgeschlossen werden. Das Argument, dass die Glaubwürdigkeitssteigerung sei, z.B. durch Zufall oder den mimetischen Isomorphismus, also einer natürlichen Verbesserung der NBE durch Anpassung an Konkurrenzunternehmen (DiMaggio & Powell, 1983, S. 151), entstanden, überzeugt nicht, da es sich um einen sprunghaften Anstieg der Glaubwürdigkeit handelt. Dennoch bedarf es weiterer Forschung, die einen längeren Zeitraum betrachtet, um diese Möglichkeit gänzlich auszuschließen. Neben der Ausweitung der Zeitreihenlänge ergibt es ebenfalls Sinn, die Querschnittseinheiten, sprich die Anzahl der beobachteten Unternehmen, zu erweitern. Gerade für Deutschland, in dem der Mittelstand den Großteil der Wirtschaftsleistung beisteuert (Röhl, 2018, S. 12), scheint die Inklusion von kleineren Unternehmen eine sinnvolle Erweiterung zu sein. Damit können die in Kapitel 3 angesprochenen Ausstrahlungseffekte auf nicht direkt durch die CSR-RL tangierte Unternehmen untersucht werden. Denkbar ist auch eine Erweiterung auf verschiedene EU-Länder, in dem verschiedene Länder miteinander hinsichtlich ihrer Glaubwürdigkeit beurteilt und verglichen werden. Bezogen auf die Stichprobengröße bleibt festzuhalten, dass eine zu geringe Beobachtungszahl vorliegt, um eine umfassende Bewertung der CSR-RL und deren Wirkung auf die Glaubwürdigkeit abschließend vorzunehmen. Eine Stichprobenvergrößerung bewirkt zusätzlich, dass die Reaktion verschiedener Branchen auf die CSR-RL besser miteinander zu vergleichen ist und die Anwendung induktiver statistischer Verfahren zulässt.

6.6. Limitationen

Eine Schwäche des verwendeten Glaubwürdigkeitsindex ist es, dass nicht alle unterschiedliche Nuancen in der NBE durch den angewandten Kodierungsrahmen erkannt werden. Beispielsweise kann die Variable Identifikation der Stakeholder unterschiedlich vom Unternehmen ausgestaltet sein. Diese können bspw. in einer aufwändigen Analyse mit grafischer Visualisierung herausgearbeitet werden oder einfach in einem Nebensatz aufgelistet sein. Obwohl beide Sachverhalte sich hinsichtlich ihrer Qualität unterscheiden, kann der Glaubwürdigkeitsindex diese Feinheiten nicht unterscheiden. Bezogen auf den Prüfungsumfang lassen manche Unternehmen nur wenige Abbildungen extern prüfen und erhalten somit ein Testat von einer Prüfungsgesellschaft. Damit werden diese Unternehmen genauso bewertet wie Unternehmen, die fast alle Inhalte extern prüfen lassen. Ferner existiert bei manchen Variablen ein Interpretationsspielraum (z.B. bei der Einbindung der Stakeholder) in der Bewertung, welcher stark vom Wortlaut in der NBE ab. Das Geschriebene muss jedoch nicht zwangsläufig mit der Realität übereinstimmen, was dazu führt, dass ggf. Sachverhalte falsch im Zuge der Inhaltsanalyse bewertet werden.

Die Bestimmung der Variablen Genauigkeit der dargestellten Informationen, Lesbarkeit und Verwendung visueller Darstellungen erfolgte nur gemäß einer kleinen Stichprobe von zufällig ausgewählten Seiten. Somit kann es zu Verzerrungen in den Ergebnissen zum Positiven aber auch zum Negativen kommen. Die Weiterentwicklung von Algorithmusbasierten Analyseverfahren zur Verarbeitung qualitativer Daten („Text Mining") kann hierbei in der Zukunft wegweisende Potentiale eröffnen.

Des Weiteren handelt es sich bei dem Glaubwürdigkeitsindex um ein akademisches Konstrukt, welches von Forschenden konzeptioniert ist (Mazzotta et al., 2020, S. 1911). Dieses Messinstrument sollte ähnlich wie in der Umfrage von Helfaya et al., 2018 durch qualitative Interviews mit Unternehmen und Stakeholdern validiert werden (Mazzotta et al., 2020, S. 1911). Damit könnte bspw. der Forschungsfrage nachgegangen werden, ob eine bestimmte Gewichtung des Glaubwürdigkeitsindex, Sinn ergibt.

In der Datenerhebung wurden nur die vorhandenen Geschäfts- und Nachhaltigkeitsberichte untersucht. Unternehmen nutzen jedoch weitere Kommunikationskanäle, wie z.B. Unternehmenswebseite, Social Media, Flyer, o.ä., um über nachhaltigkeitsbezogene Aspekte zu berichten (Seele & Lock, 2015, S. 406). Ergo kann es sein, dass publizierte NFI

nicht in die Inhaltsanalyse eingeflossen sind. Das Auslassen von NFI kann zu einer schlechteren Bewertung führen.

Nach Krippendorff (2004, S. 214-215) sind Daten aus einer Inhaltsanalyse nur dann zuverlässig, wenn die Kriterien Stabilität sowie Replizierbarkeit erfüllt sind. Das Kriterium der Stabilität besagt, dass unabhängig vom Zeitpunkt der Kodierung immer das gleiche Resultat erfolgen muss (Krippendorff, 2004, S. 215). Im Rahmen der Masterarbeit kann die Stabilität ausreichend kontrolliert werden, sodass intrasubjektive Inkonsistenzen unwahrscheinlich sind. Trotzdem können menschliche Fehler im Rahmen der Kodierung nicht vollständig ausgeschlossen werden (z.B. das Verrutschen in der Zeile o.ä.). Zudem ist an dieser Stelle anzumerken, dass für eine Inhaltsanalyse die Vorgehensweise der Kodierung akribisch dokumentiert sein muss, da ansonsten die Kodierung stets unterschiedlich ausfällt. Ferner besagt die Replizierbarkeit, dass verschiedene Anwender mittels der vorgebenden Kodierung zum gleichen Ergebnis kommen müssen (Krippendorff, 2004, S. 215). Beispielsweise berechnet Krippendorff's Alpha, wie hoch die Korrelation zwischen den intersubjektiven Bewertungen verschiedener Anwender ist (Krippendorff, 2004, S. 1-2). Aufgrund mangelnder personeller Ressourcen konnte dieses Kriterium der Reliabilität nicht geprüft werden. In der Forschungsarbeit von Mazzotta et al. (2020) ist nicht bei jeder Variable eindeutig beschrieben, wie die Kodierung genau erfolgt ist. Aus diesem Grund musste teils auf alternative Bewertungen zurückgegriffen werden. Dies zeigt die Relevanz der exakten Erläuterung des Kodierungsrahmens im Kontext einer Inhaltsanalyse.

7. Fazit

Das Instrumentarium der NBE konnte im Rahmen der Masterarbeit als wichtiges Kommunikationsmedium zwischen dem Unternehmen und seiner Umwelt herausgearbeitet werden. Der Zusammenhang zwischen einer glaubwürdigen NBE und Legitimität konnte mit Hilfe der bestehenden Literatur hergestellt werden. Die Synthese aus dem Vier-Stufen-Pyramiden-Modell sowie der Legitimitätstheorie prognostizieren eine positive Auswirkung der CSR-RL auf die Glaubwürdigkeit der NBE. Ein Unternehmen strebt stets nach einem möglichst hohen Maß an Legitimität. Aufgrund der pragmatischen Legitimität ist davon auszugehen, dass ein Unternehmen aus Eigeninteresse (z.B. Vermeidung von Strafzahlungen durch die Nichteinhaltung des Gesetzes) den neuen Vorgaben der CSR-RL Folge leistet. Ferner wurde aufgezeigt, dass die Erwartungshaltung der Stakeholder an die ethische und moralische Verantwortung eines Unternehmens gestiegen ist. Nur in einem deliberativen Diskurs mit seinen Stakeholdern kann moralische Legitimität erzeugt werden. Dieser Austausch, der u.a. über die NBE erfolgt, sollte demnach glaubwürdig sein, da die Glaubwürdigkeit als Voraussetzung für Legitimität betrachtet werden kann. Da durch die CSR-RL Indikatoren für eine glaubwürdige NBE tangiert werden (z.B. Vorgabe der Mindestinhalte), ist von einer Glaubwürdigkeitssteigerung in der Berichterstattung auszugehen.

Im Rahmen der Inhaltsanalyse, in der 41 deutsche Unternehmen des MDAX untersucht wurden, können folgende Kernergebnisse subsumiert werden: Die Einführung der CSR-RL hat dazu geführt, dass acht Unternehmen, die zuvor keine NBE hatten, zur Veröffentlichung von NFI gezwungen waren. Damit ist es dem Adressatenkreis der CSR-RL nicht mehr möglich, auf die Informationspreisgabe von NFI zu verzichten. Alleine durch diesen Effekt ist insgesamt die Glaubwürdigkeit angestiegen. Darüber hinaus hat sich die Glaubwürdigkeit der NBE nach Einführung der CSR-RL verbessert. Dieser Anstieg wird insbesondere durch die Dimensionen Wahrheit und Wahrhaftigkeit initiiert. Hingegen hat die CSR-RL keine Auswirkung auf die Dimension Angemessenheit & Verständlichkeit gezeigt. Im Rahmen der multiplen linearen Regressionsmodelle konnte für mögliche Störeinflüsse kontrolliert werden. Die Resultate belegen, dass die Einflussgrößen Unternehmensgröße, Eigentumsstreuung sowie Umweltsensitivität die Glaubwürdigkeit der NBE positiv beeinflussen. Zusätzlich gibt es Indizien, die dafür sprechen, dass die Vermögens- und Ertragslage ebenso einen positiven Einfluss auf die Glaubwürdigkeit der NBE ausübt.

Wenngleich die verwendete Forschungsmethodik nicht frei von Limitationen ist, liefert die Masterarbeit zusätzliche empirische Evidenz dafür, dass eine verpflichtende NBE zu einer Verbesserung der Glaubwürdigkeit führt. Damit konnte das Ziel der EU, die Beziehung zwischen Unternehmen und den Stakeholdern zu verbessern, erreicht werden. Nichtsdestotrotz gibt es bei vielen Unternehmen Verbesserungspotential in der NBE. Zum Beispiel schaffen es viele Unternehmen noch nicht, ihre Stakeholder aktiv in Entscheidungsprozesse einzubinden. Um aussagekräftige Vergleiche mit anderen EU-Ländern oder unterschiedlichen Gesetzgebungen herzustellen, bedarf es weiterer Forschung.

Literatur

Ackers, B. (2009). Corporate social responsibility assurance: How do south african publicly listed companies compare? [Journal Article]. *Meditari Accountancy Research*, *17* (2), 1-17.

Adams, C. A. & Evans, R. (2004). Accountability, completeness, credibility and the audit expectations gap [Journal Article]. *Journal of Corporate Citizenship*, *14*, 97-115.

Adaui, C. R. L. (2020). Sustainability reporting quality of peruvian listed companies and the impact of regulatory requirements of sustainability disclosures [Journal Article]. *Sustainability*, *12* (3), 1135.

Aggarwal, P. (2013). Sustainability reporting and its impact on corporate financial performance: A literature review [Journal Article]. *Indian Journal of Commerce and Management Studies*, *4* (3), 51-59.

Aifuwa, H. O. (2020). Sustainability reporting and firm performance in developing climes: A review of literature [Journal Article]. *Copernican Journal of Finance & Accounting*, *9* (1), 9-29.

Ali, W., Frynas, J. G. & Mahmood, Z. (2017). Determinants of corporate social responsibility (csr) disclosure in developed and developing countries: A literature review [Journal Article]. *Corporate Social Responsibility and Environmental Management*, *24* (4), 273-294.

Almeling, C., Flick, C. & Scharr, C. (2020). *Abschlussprüfung klipp & klar* [Book]. Springer.

Amel-Zadeh, A. & Serafeim, G. (2018). Why and how investors use esg information: Evidence from a global survey [Journal Article]. *Financial Analysts Journal*, *74* (3), 87-103.

Amran, A., Lee, S. P. & Devi, S. S. (2014). The influence of governance structure and strategic corporate social responsibility toward sustainability reporting quality [Journal Article]. *Business Strategy and the Environment*, *23* (4), 217-235.

Ankele, K. & Grothe, A. (2019). Strategisches nachhaltigkeitsmanagement durch nachhaltigkeitsbewertung [Book Section]. In *Nachhaltiges management* (S. 551-574). Springer.

Arena, C., Bozzolan, S. & Michelon, G. (2015). Environmental reporting: Transparency to stakeholders or stakeholder manipulation? an analysis of disclosure tone and the role of the board of directors [Journal Article]. *Corporate Social Responsibility and Environmental Management*, *22* (6), 346-361.

Arnstein, S. R. (1969). A ladder of citizen participation [Journal Article]. *Journal of the American Institute of planners*, *35* (4), 216-224.

Ayuso, S., Roca, M., Arevalo, J. A. & Aravind, D. (2016). What determines principle-based standards implementation? reporting on global compact adoption in spanish firms [Journal Article]. *Journal of business ethics*, *133* (3), 553-565.

Böcking, H.-J. & Althoff, C. (2017). Paradigmenwechsel in der (konzern-) lageberichterstattung über nicht-monetäre erfolgsfaktoren: Prefinancial performance indicators als vorstufe, nicht als gegensatz von financial performance indicators [Journal Article]. *Der Konzern*, *15* (5), 246-255.

Bentele, G. & Nothhaft, H. (2011). Vertrauen und glaubwürdigkeit als grundlage von corporate social responsibility: Die (massen-) mediale konstruktion von verantwortung und verantwortlichkeit [Book Section]. In *Handbuch csr* (S. 45-70). Springer.

Beretta, S. & Bozzolan, S. (2008). Quality versus quantity: The case of forward-looking disclosure [Journal Article]. *Journal of Accounting, Auditing & Finance*, *23* (3), 333-376.

Berger, P. L. & Luckmann, T. (2004). *Die gesellschaftliche konstruktion der wirklichkeit* [Book]. Verlag Fischer.

Berrone, P., Ricart, J. E., Duch, A. I., Bernardo, V., Salvador, J., Piedra Peña, J. & Rodríguez Planas, M. (2019). Easier: An evaluation model for public–private partnerships contributing to the sustainable development goals [Journal Article]. *Sustainability*, *11* (8), 2339.

Berthelot, S., Coulmont, M. & Serret, V. (2012). Do investors value sustainability reports? a canadian study [Journal Article]. *Corporate Social Responsibility and Environmental Management*, *19* (6), 355-363.

Bewley, K. & Li, Y. (2000). Disclosure of environmental information by canadian manufacturing companies: a voluntary disclosure perspective [Book Section]. In *Advances in environmental accounting & management* (S. 201-226). Emerald Group Publishing Limited.

Biswas, P. K., Mansi, M. & Pandey, R. (2018). Board composition, sustainability committee and corporate social and environmental performance in australia [Journal Article]. *Pacific Accounting Review*, *30* (4), 517-540.

Blackwelder, B., Coleman, K., Colunga-Santoyo, S., Harrison, J. S. & Wozniak, D. (2016). *The volkswagen scandal* [Book]. University of Richmond.

Boiral, O. (2013). Sustainability reports as simulacra? a counter-account of a and a+ gri reports [Journal Article]. *Accounting, Auditing & Accountability Journal*, *26* (7), 1036-1071.

Brammer, S. & Millington, A. (2006). Firm size, organizational visibility and corporate philanthropy: An empirical analysis [Journal Article]. *Business Ethics: A European Review*, *15* (1), 6-18.

Brammer, S. & Pavelin, S. (2006). Voluntary environmental disclosures by large uk companies [Journal Article]. *Journal of Business Finance & Accounting*, *33* (7-8), 1168-1188.

Brammer, S. & Pavelin, S. (2008). Factors influencing the quality of corporate environmental disclosure [Journal Article]. *Business strategy and the environment*, *17* (2), 120-136.

Brundtland, G. H. (1987). *Report of the world commission on eviornment and development: Our commom future* (Report). Zugriff auf https://www.are.admin.ch/are/de/home/nachhaltige-entwicklung/internationale-zusammenarbeit/agenda2030/uno-_-meilensteine-zur-nachhaltigen-entwicklung/1987-brundtland-bericht.html

Bundesregierung. (2017). (Web Page Nr. 03.02.2021). Zugriff auf https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/RegE_CSR-Richtlinie.pdf?__blob=publicationFile&v=1

Carroll, A. B. (1979). A three-dimensional conceptual model of corporate performance [Journal Article]. *Academy of management review*, *4* (4), 497-505.

Carroll, A. B. (1991). The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders [Journal Article]. *Business horizons*, *34* (4), 39-48.

Carroll, A. B. (2016). Carroll's pyramid of csr: Taking another look [Journal Article]. *International journal of corporate social responsibility*, *1* (1), 1-8.

Carroll, A. B. & Buchholtz, A. K. (2006). *Ethics and stakeholder management* [Book]. Thomson South-Western.

Carroll, A. B., Lipartito, K. J., Post, J. E. & Werhane, P. H. (2012). *Corporate responsibility: The american experience* [Book]. Cambridge University Press.

Carrots & Sticks. (2016). *Global trends in sustainability reporting - regulation and policy* (Report). Zugriff auf https://assets.kpmg/content/dam/kpmg/pdf/2016/05/carrots-and-sticks-may-2016.pdf

Chau, G. K. & Gray, S. J. (2002). Ownership structure and corporate voluntary disclosure in hong kong and singapore [Journal Article]. *The International Journal of Accounting*, *37* (2), 247-265.

Cho, C. H., Laine, M., Roberts, R. W. & Rodrigue, M. (2015). Organized hypocrisy, organizational façades, and sustainability reporting [Journal Article]. *Accounting, organizations and society*, *40*, 78-94.

Clarke, T. (2007). The materiality of sustainability: Corporate social and environmental responsibility as instruments of strategic change? [Journal Article]. *Corporate Governance and Sustainability: Challenges for Theory and Practice*, *1*, 219-251.

Cormier, D., Ledoux, M., Magnan, M. & Aerts, W. (2010). Corporate governance and information asymmetry between managers and investors [Journal Article]. *Corporate Governance: The international journal of business in society*, *10* (5), 574-589.

Cormier, D. & Magnan, M. (1999). Corporate environmental disclosure strategies: Determinants, costs and benefits [Journal Article]. *Journal of Accounting, Auditing & Finance*, *14* (4), 429-451.

Cormier, D. & Magnan, M. (2003). Environmental reporting management: A continental european perspective [Journal Article]. *Journal of Accounting and public Policy*, *22* (1), 43-62.

Cormier, D., Magnan, M. & Van Velthoven, B. (2005). Environmental disclosure quality in large german companies: economic incentives, public pressures or institutional conditions? [Journal Article]. *European accounting review*, *14* (1), 3-39.

Cucari, N., Esposito De Falco, S. & Orlando, B. (2018). Diversity of board of directors and environmental social governance: Evidence from italian listed companies [Journal Article]. *Corporate Social Responsibility*

and Environmental Management, *25* (3), 250-266.

Daub, C.-H. (2010). Gesellschaftliche legitimation durch glaubwürdige berichterstattung [Journal Article]. *Ökologisches Wirtschaften-Fachzeitschrift*, *25* (1), 30-34.

DAX-Indices. (2021, 01.03.2021). (Web Page Nr. 25.03.2021). Zugriff auf https://www.dax-indices.com/document/Resources/Guides/Historical_Index_Compositions.pdf

Deegan, C. & Gordon, B. (1996). A study of the environmental disclosure practices of australian corporations [Journal Article]. *Accounting and business research*, *26* (3), 187-199.

Demuijnck, G. & Fasterling, B. (2016). The social license to operate [Journal Article]. *Journal of business ethics*, *136* (4), 675-685.

Dienes, D., Sassen, R. & Fischer, J. (2016). What are the drivers of sustainability reporting? a systematic review [Journal Article]. *Sustainability Accounting, Management and Policy Journal*, *7* (2), 154-189.

Dilling, P. F. (2010). Sustainability reporting in a global context: What are the characteristics of corporations that provide high quality sustainability reports an empirical analysis [Journal Article]. *International Business & Economics Research Journal (IBER)*, *9* (1), 19-30.

DiMaggio, P. J. & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields [Journal Article]. *American sociological review*, 147-160.

DNK. (2018). *Der dnk im sinne des csr-richtlinie-umsetzungsgesetzes (csr-rug)–eine orientierungshilfe für anwender* (Report). Zugriff auf https://www.deutscher-nachhaltigkeitskodex.de/de-DE/Documents/PDFs/Sustainability-Code/DNK-Orientierungshilfe-Anwender-2018-de.aspx

Dong, S. & Xu, L. (2016). The impact of explicit csr regulation: Evidence from china's mining firms [Journal Article]. *Journal of Applied Accounting Research*, *17* (2), 237-258.

Dumay, J., Frost, G. & Beck, C. (2015). Material legitimacy [Journal Article]. *Journal of Accounting & Organizational Change*, *11* (1), 2-23.

Dumitru, M., Dyduch, J., Guşe, R.-G. & Krasodomska, J. (2017). Corporate reporting practices in poland and romania–an ex-ante study to the new non-financial reporting european directive [Journal Article]. *Accounting in Europe*, *14* (3), 279-304.

Eccles, R. G. & Krzus, M. P. (2010). *One report: Integrated reporting for a sustainable strategy* [Book]. John Wiley & Sons.

Eccles, R. G. & Saltzman, D. (2011). Achieving sustainability through integrated reporting [Journal Article]. *Stanford Social Innovation Review*, *9* (3), 56-61.

Edgley, C., Jones, M. J. & Atkins, J. (2015). The adoption of the materiality concept in social and environmental reporting assurance: A field study approach [Journal Article]. *The British Accounting Review*, *47* (1), 1-18.

Eisend, M. (2003). *Glaubwürdigkeit in der marketingkommunikation: Konzeption, einflussfaktoren und wirkungspotenzial* [Book]. Springer-Verlag.

Elkington, J. (1997). The triple bottom line [Book Section]. In M. V. Russo (Hrsg.), *Environmental management: Readings and cases* (S. 49-68). Sage Publications.

EU. (2014a, 15.04.2014). (Web Page Nr. 20.04.2021). Zugriff auf https://ec.europa.eu/commission/presscorner/detail/cs/MEMO_14_301

EU. (2014b, 15.11.2014). (Web Page Nr. 25.01.2021). Zugriff auf https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=celex%3A32014L0095

Farooq, M. B. & De Villiers, C. (2017). The market for sustainability assurance services [Journal Article]. *Pacific Accounting Review*, *29* (1), 79-106.

Ferrero, I., Michael Hoffman, W. & McNulty, R. E. (2014). Must milton friedman embrace stakeholder theory? [Journal Article]. *Business and Society Review*, *119* (1), 37-59.

Fifka, M. (2014). *Csr und reporting* [Book]. Springer.

Fifka, M. S. & Drabble, M. (2012). Focus and standardization of sustainability reporting–a comparative study of the united kingdom and finland [Journal Article]. *Business Strategy and the Environment*, *21* (7), 455-474.

Flesch, R. (1948). A new readability yardstick [Journal Article]. *Journal of applied psychology*, *32* (3), 221-233.

Frankel, R., McNichols, M. & Wilson, G. P. (1995). Discretionary disclosure and external financing [Journal Article]. *Accounting Review*, 135-150.

Frias-Aceituno, J. V., Rodríguez-Ariza, L. & García-Sánchez, I. M. (2013). Is integrated reporting determined by a country's legal system? an exploratory study [Journal Article]. *Journal of cleaner production*, *44*, 45-55.

Friedman, M. (2007). The social responsibility of business is to increase its profits [Book Section]. In W. C. Zimmerli, M. Holzinger & R. K. (Hrsg.), *Corporate ethics and corporate governance* (S. 173-178). Springer.

Gallego-Álvarez, I. & Ortas, E. (2017). Corporate environmental sustainability reporting in the context of national cultures: A quantile regression approach [Journal Article]. *International Business Review*, *26* (2), 337-353.

Gamerschlag, R., Möller, K. & Verbeeten, F. (2011). Determinants of voluntary csr disclosure: Empirical evidence from germany [Journal Article]. *Review of Managerial Science*, *5* (2-3), 233-262.

Gillet, C. (2012). A study of sustainability verification practices: The french case [Journal Article]. *Journal of Accounting & Organizational Change*, *8* (1), 62-84.

Gray, R. (2000). Current developments and trends in social and environmental auditing, reporting and attestation: a review and comment [Journal Article]. *International journal of auditing*, *4* (3), 247-268.

GRI. (2016). (Web Page Nr. 11.01.2021). Zugriff auf https://www.globalreporting.org/how-to-use-the-gri-standards/gri-standards-english-language/

Grob, L., Sydow, J. & Heinz, R. (2020). *Bewertung der csr-richtlinie und ihrer umsetzung in deutschen konzernen der rohstoffverarbeitenden industrie (policy paper)* (Report). Zugriff auf https://germanwatch.org/sites/default/files/CSR_Bericht_1.pdf

Gulenko, M. (2018). Mandatory csr reporting—literature review and future developments in germany [Journal Article]. *NachhaltigkeitsManagementForum| Sustainability Management Forum*, *26*, 3-17.

Habermas, J. (2002). *On the pragmatics of social interaction: Preliminary studies in the theory of communicative action* [Book]. MIT Press.

Hagest, J. (1975). *Zur logik der prüferischen Überzeugungsbildung bei jahresabschlußprüfungen* [Book]. Ludwig-Maximiliams-Universität.

Hahn, R. & Kühnen, M. (2013). Determinants of sustainability reporting: A review of results, trends, theory, and opportunities in an expanding field of research [Journal Article]. *Journal of cleaner production*, *59*, 5-21.

Haji, A. A. (2013). Corporate social responsibility disclosures over time: Evidence from malaysia [Journal Article]. *Managerial auditing journal*.

Hąbek, P. & Wolniak, R. (2016). Assessing the quality of corporate social responsibility reports: The case of reporting practices in selected european union member states [Journal Article]. *Quality & quantity*, *50* (1), 399-420.

Heckler, A. & Bröcker, N. (2017). Die csr-berichtspflicht in der hauptversammlungssaison 2018 [Journal Article]. *Die Aktiengesellschaft* (21), 761-770.

Helfaya, A. & Moussa, T. (2017). Do board's corporate social responsibility strategy and orientation influence environmental sustainability disclosure? uk evidence [Journal Article]. *Business Strategy and the Environment*, *26* (8), 1061-1077.

Helfaya, A. & Whittington, M. (2019). Does designing environmental sustainability disclosure quality measures make a difference? [Journal Article]. *Business Strategy and the Environment*, *28* (4), 525-541.

Helfaya, A., Whittington, M. & Alawattage, C. (2018). Exploring the quality of corporate environmental reporting: Surveying preparers' and users' perceptions [Journal Article]. *Accounting, Auditing & Accountability Journal*, *31* (1), 163-193.

Hodge, K., Subramaniam, N. & Stewart, J. (2009). Assurance of sustainability reports: Impact on report users' confidence and perceptions of information credibility [Journal Article]. *Australian accounting review*, *19* (3), 178-194.

Hooks, J. & van Staden, C. J. (2011). Evaluating environmental disclosures: The relationship between quality and extent measures [Journal Article]. *The British Accounting Review*, *43* (3), 200-213.

Hoozée, S., Maussen, S. & Vangronsveld, P. (2019). The impact of readability of corporate social responsibility information on credibility as perceived by generalist versus specialist readers [Journal Artic-

le]. *Sustainability Accounting, Management and Policy Journal*, *10* (3), 570-591.

ICC. (2015, 12.04.2021). *Sustainability reporting - future directions (policy statement)* (Report). Zugriff auf https://iccwbo.org/content/uploads/sites/3/2016/10/Sustainability-Reporting-Future-directions.pdf

Jones, P., Comfort, D. & Hillier, D. (2016). Materiality in corporate sustainability reporting within uk retailing [Journal Article]. *Journal of Public Affairs*, *16* (1), 81-90.

Kajüter, P. (2017a). Das csr-richtlinie-umsetzungsgesetz–ein kompromiss [Journal Article]. *Zeitschrift für internationale Rechnungslegung*, *12* (4), 137-138.

Kajüter, P. (2017b). Nichtfinanzielle berichterstattung nach dem csr-richtlinie-umsetzungsgesetz [Journal Article]. *Der Betrieb*, *70* (12), 617-624.

Kelly, J. D. (1993). The effects of display format and data density on time spent reading statistics in text, tables and graphs [Journal Article]. *Journalism Quarterly*, *70* (1), 140-149.

Kerret, D., Menahem, G. & Sagi, R. (2010). Effects of the design of environmental disclosure regulation on information provision: The case of israeli securities regulation [Journal Article]. *Environmental Science & Technology*, *44* (21), 8022-8029.

Köhnken, G. (1990). Glaubwürdigkeit: Untersuchungen zu einem psychologischen konstrukt [Journal Article]. *Psychologie-Verlag-Union*.

Kolk, A. & Pinkse, J. (2010). The integration of corporate governance in corporate social responsibility disclosures [Journal Article]. *Corporate social responsibility and environmental management*, *17* (1), 15-26.

Krasodomska, J., Simnett, R. & Street, D. L. (2021). Extended external reporting assurance: Current practices and challenges [Journal Article]. *Journal of International Financial Management & Accounting*, *32*, 104-142.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* [Book]. London, Neu-Delhi: Sage Publications.

Kropp, A. (2018). *Grundlagen der nachhaltigen entwicklung: Handlungsmöglichkeiten und strategien zur umsetzung* [Book]. Wiesbaden: Springer-Verlag.

Lackmann, J. (2010). Leitlinien zur nachhaltigkeitsberichterstattung [Book Section]. In *Die auswirkungen der nachhaltigkeitsberichterstattung auf den kapitalmarkt* (S. 39-89). Springer.

La Torre, M., Sabelfeld, S., Blomkvist, M., Tarquinio, L. & Dumay, J. (2018). Harmonising non-financial reporting regulation in europe: Practical forces and projections for future research [Journal Article]. *Meditari Accountancy Research*, *26* (4), 598-621.

Leffson, U., Lippmann, K. & Baetge, J. (1969). *Zur sicherheit und wirtschaftlichkeit der urteilsbildung bei prüfungen* [Book]. Verlagsbuchhandlung des Instituts der Wirtschaftsprüfer.

Li, Y. (2010). The case analysis of the scandal of enron [Journal Article]. *International Journal of business and management*, *5* (10), 37-41.

Liao, L., Luo, L. & Tang, Q. (2015). Gender diversity, board independence, environmental committee and greenhouse gas disclosure [Journal Article]. *The British Accounting Review*, *47* (4), 409-424.

Lock, I. (2016). Glaubwürdigkeit in der csr-kommunikation [Journal Article]. *Publizistik*, *61* (4), 413-429.

Lock, I. & Schulz-Knappe, C. (2019). Credible corporate social responsibility (csr) communication predicts legitimacy [Journal Article]. *Corporate Communications: An International Journal*, *24* (1), 2-20.

Lock, I. & Seele, P. (2016). The credibility of csr (corporate social responsibility) reports in europe. evidence from a quantitative content analysis in 11 countries [Journal Article]. *Journal of Cleaner Production*, *122*, 186-200.

Loew, T., Ankele, K., Braun, S. & Clausen, J. (2004). *Bedeutung der internationalen csr-diskussion für nachhaltigkeit und die sich daraus ergebenden anforderungen an unternehmen mit fokus berichterstattung* (Report). Zugriff auf https://www.upj.de/fileadmin/user_upload/MAIN-dateien/Themen/Einfuehrung/ioew_csr_diskussion_2004.pdf

Lubitzsch, K. (2008). *Prüfungssicherheit bei betriebswirtschaftlichen prüfungen: eine theoretische und empirische analyse* [Book]. IDW-Verlag.

Maas, K. & Sampers, P. (2020). The expected impacts of regulating non-financial reporting [Journal Article]. *Maandblad Voor Accountancy en Bedrijfseconomie*, *94* (7/8), 265-274.

Manetti, G. (2011). The quality of stakeholder engagement in sustainability reporting: empirical evidence and critical points [Journal Article]. *Corporate social responsibility and environmental management*, *18* (2), 110-122.

Manetti, G. & Becatti, L. (2009). Assurance services for sustainability reports: Standards and empirical evidence [Journal Article]. *Journal of Business Ethics*, *87* (1), 289-298.

Marten, K., Quick, R. & Ruhnke, K. (2007). *Wirtschaftsprüfung* [Book]. Schäffer-Poeschel Verlag.

Mazzotta, R., Bronzetti, G. & Veltri, S. (2020). Are mandatory non-financial disclosures credible? evidence from italian listed companies [Journal Article]. *Corporate Social Responsibility and Environmental Management*, *27* (4), 1900-1913.

Meek, G. K., Roberts, C. B. & Gray, S. J. (1995). Factors influencing voluntary annual report disclosures by us, uk and continental european multinational corporations [Journal Article]. *Journal of international business studies*, *26* (3), 555-572.

Merkt, H. (2001). *Unternehmenspublizität: Offenlegung von unternehmensdaten als korrelat der marktteilnahme* [Book]. Mohr Siebeck.

Michalczuk, G. & Konarzewska, U. (2020). Standardization of corporate social responsibility reporting using the gri framework [Journal Article]. *Optimum. Economic Studies*, *1* (99), 74-88.

Michelon, G. & Parbonetti, A. (2012). The effect of corporate governance on sustainability disclosure [Journal Article]. *Journal of management & governance*, *16* (3), 477-509.

Michelon, G., Pilonato, S. & Ricceri, F. (2015). Csr reporting practices and the quality of disclosure: An empirical analysis [Journal Article]. *Critical perspectives on accounting*, *33*, 59-78.

Mio, C., Panfilo, S. & Blundo, B. (2020). Sustainable development goals and the strategic role of business: A systematic literature review [Journal Article]. *Business Strategy and the Environment*, *29* (8), 3220-3245.

Mion, G. & Loza Adaui, C. R. (2019). Mandatory nonfinancial disclosure and its consequences on the sustainability reporting quality of italian and german companies [Journal Article]. *Sustainability*, *11* (17), 1-28.

Moore, J. E., Mascarenhas, A., Bain, J. & Straus, S. E. (2017). Developing a comprehensive definition of sustainability [Journal Article]. *Implementation Science*, *12* (1), 1-8.

Neu, D., Warsame, H. & Pedwell, K. (1998). Managing public impressions: environmental disclosures in annual reports [Journal Article]. *Accounting, organizations and society*, *23* (3), 265-282.

Perego, P. & Kolk, A. (2012). Multinationals' accountability on sustainability: The evolution of third-party assurance of sustainability reports [Journal Article]. *Journal of Business Ethics*, *110* (2), 173-190.

Peters, G. F. & Romi, A. M. (2015). The association between sustainability governance characteristics and the assurance of corporate sustainability reports [Journal Article]. *Auditing: A Journal of Practice & Theory*, *34* (1), 163-198.

Pufé, I. (2017). *Nachhaltigkeit* [Book]. UTB.

QONTIGO. (2021, 20.04.2021). (Web Page Nr. 20.04.2021). Zugriff auf https://www.dax-indices.com/index-details?isin=DE0008467416

Rehbinder, E. (2014). Corporate social responsibility–von der gesellschaftspolitischen forderung zur rechtlichen verankerung [Book Section]. In S. Deinert, C. Schrader & B. Stoll (Hrsg.), *Corporate social responsibility (csr). die richtlinie 2014/95/eu - chancen und herausforderungen* (S. 10-37). Kassel University Press.

Rennekamp, K. (2012). Processing fluency and investors' reactions to disclosure readability [Journal Article]. *Journal of Accounting Research*, *50* (5), 1319-1354.

Reynolds, M. & Yuthas, K. (2008). Moral discourse and corporate social responsibility reporting [Journal Article]. *Journal of business ethics*, *78* (1), 47-64.

Röhl, K.-H. (2018). *Mid caps: der große mittelstand. die wirtschaftspolitik berücksichtigt die relevanz der mid caps nicht ausreichend* (Report). Zugriff auf https://www.iwkoeln.de/fileadmin/user_upload/Studien/policy_papers/PDF/2018/IW-Policy_Paper_2018_4_Mid_Caps.pdf

Roca, L. C. & Searcy, C. (2012). An analysis of indicators disclosed in corporate sustainability reports [Journal Article]. *Journal of cleaner production*, *20* (1), 103-118.

Romero, S., Fernandez-Feijoo, B. & Ruiz, S. (2014). Perceptions of quality

of assurance statements for sustainability reports [Journal Article]. *Social Responsibility Journal*, *10* (3), 480-499.

Rupley, K. H., Brown, D. & Marshall, R. S. (2012). Governance, media and the quality of environmental disclosure [Journal Article]. *Journal of Accounting and Public Policy*, *31* (6), 610-640.

Sachs, J. D. (2012). From millennium development goals to sustainable development goals [Journal Article]. *The Lancet*, *379* (9832), 2206-2211.

Sandhu, S. (2012). *Public relations und legitimität: Der beitrag des organisationalen neo-institutionalismus für die pr-forschung* [Book]. Springer-Verlag.

Schreier, M. (2014). Varianten qualitativer inhaltsanalyse: Ein wegweiser im dickdicht der begrifflichkeiten [Journal Article]. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, *15* (1), 1-27.

Schwarze, J. (2011). Soft law im recht der europäischen union [Journal Article]. *EuR Europarecht*, *46* (1), 3-18.

Schweren, F. C. & Brink, A. (2016). Csr-berichterstattung in europa [Journal Article]. *zfwu Zeitschrift für Wirtschafts-und Unternehmensethik*, *17* (1), 177-191.

SEC. (1998). *A plain english handbook: How to create clear sec disclosure documents* [Book]. The Office.

Seele, P. & Lock, I. (2015). Instrumental and/or deliberative? a typology of csr communication tools [Journal Article]. *Journal of Business Ethics*, *131* (2), 401-414.

Siano, A., Vollero, A., Conte, F. & Amabile, S. (2017). "more than words": Expanding the taxonomy of greenwashing after the volkswagen scandal [Journal Article]. *Journal of Business Research*, *71*, 27-37.

Siew, R. Y. (2015). A review of corporate sustainability reporting tools (srts) [Journal Article]. *Journal of environmental management*, *164*, 180-195.

Smith, J., Haniffa, R. & Fairbrass, J. (2011). A conceptual framework for investigating 'capture'in corporate sustainability reporting assurance [Journal Article]. *Journal of Business Ethics*, *99* (3), 425-439.

Smith, M. & Taffler, R. J. (2000). The chairman's statement-a content analysis of discretionary narrative disclosures [Journal Article]. *Accounting, Auditing & Accountability Journal*, *13* (5), 624-647.

Spelthahn, S., Fuchs, L. & Demele, U. (2009). Glaubwürdigkeit in der nachhaltigkeitsberichterstattung [Journal Article]. *uwf UmweltWirtschaftsForum*, *17* (1), 61-68.

Stawinoga, M. (2017). Die richtlinie 2014/95/eu und das csr-richtlinie-umsetzungsgesetz–eine normative analyse des transformationsprozesses sowie daraus resultierender implikationen für die rechnungslegungs-und prüfungspraxis [Journal Article]. *uwf UmweltWirtschaftsForum| Sustainability Management Forum*, *25*, 213-227.

Stibi, E.-M. (1995). *Prüfungsrisikomodell und risikoorientierte abschlußprüfung* [Book]. IDW-Verlag.

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches [Journal Article]. *Academy of management review*, *20* (3), 571-610.

Szabó, D. G. & Sørensen, K. E. (2015). New eu directive on the disclosure of non-financial information (csr) [Journal Article]. *European Financial Law Review*, *12*, 307-340.

Tan, H., YING WANG, E. & Zhou, B. (2014). When the use of positive language backfires: The joint effect of tone, readability, and investor sophistication on earnings judgments [Journal Article]. *Journal of Accounting Research*, *52* (1), 273-302.

Waddock, S. & Goggins, B. K. (2011). The paradoxes of communicating corporate social responsibility [Book Section]. In . S. M. O. Ihlen J. L. Bartlett (Hrsg.), *The handbook of communication and corporate social repsonsibility* (S. 23-43). John Wiley & Sons.

Walden, D. & Depping, A. (2015). *Csr und recht* [Book]. Springer.

Walterbusch, M., Handzlik, J. & Teuteberg, F. (2013). Status quo der wirtschaftsprüfung von nachhaltigkeitsberichten [Book Section]. In *It-gestütztes ressourcen-und energiemanagement* (S. 309-321). Springer.

Wiseman, J. (1982). An evaluation of environmental disclosures made in corporate annual reports [Journal Article]. *Accounting, organizations and society*, *7* (1), 53-63.

Wong, R. & Millington, A. (2014). Corporate social disclosures: A user perspective on assurance [Journal Article]. *Accounting, Auditing & Accountability Journal*, *27* (5), 863-887.

Yunus, S., Elijido-Ten, E. & Abhayawansa, S. (2016). Determinants of carbon management strategy adoption: Evidence from australia's top 200 publicly listed firms [Journal Article]. *Managerial Auditing Journal*, *31* (2), 156-179.

Zwick, Y. (2018). Der deutsche nachhaltigkeitskodex [Book Section]. In *Csr und fashion* (S. 63-77). Springer.

Junior Management Science 7(3) (2022) 643-667

Junior Management Science

journal homepage: www.jums.academy

Demand Estimation for Solar Photovoltaics in the United States – An Instrumental Variable Approach

Veronique Clara Bukow

Technische Universität München

Abstract

Worldwide the demand for solar photovoltaics (PV) has increased significantly over the past decades. This was driven by a price reduction for solar PV systems. A two-stage least squares linear regression yields insights into the price sensitivity for residential customers in the U.S., and California in particular. The specification includes instrumental variables as well as fixed effects to account for the common issues of endogeneity and data heterogeneity in demand estimation problems, respectively. The variation in the sales tax rate on solar PV and the movements of polysilicon spot prices are used to instrumentalise PV price changes. The regression results imply an inelastic demand with a long-term price elasticity of -0.443, accounting for differences over state and time. Investigating price elasticities for various income groups shows that lower-income customers react more strongly to price changes compared to those with relatively high income (-0.521 vs. -0.195). Likewise, regions with lower population density are more sensitive to price changes (-0.473 vs. -0.338). Besides price, installation costs and technological efficiency majorly impact the system size installed. Results of this study can provide data-driven guidance to efficient policy design and pricing strategies.

Keywords: Price elasticity; solar photovoltaic; instrumental variables; demand estimation.

1. Introduction

1.1. Central Issue

> *"The path towards sustainable energy sources will be long and sometimes difficult. But America cannot resist this transition; we must lead it."*
>
> *Barack Obama, 2013*

Mounting greenhouse gas emissions (GHG) and global climate change put especially industrialized countries under pressure to act. In an attempt to reduce emissions and limit global warming, the development and deployment of renewable energy sources is increasing worldwide as well as in the United States (U.S.) (EIA, 2019c; IEA, 2019). Hence, the importance of and focus on solar photovoltaics (PV) as one source of renewable energy has risen continuously in the past years. According to the International Energy Agency (IEA), the technology is expected to be the main accelerator in renewable capacity growth from 2019 to 2024 (IEA, 2019). It is little surprising, therefore, that investors, governments, and researchers are taking interest in understanding which factors predominantly drive the solar PV demand. Especially in the U.S., one of the world's largest economies with many different regional characteristics and incentive policies, this issue is of vital importance for both firms and policymakers (Gillingham & Tsvetanov, 2019a). Naturally, the installation price and subsequent maintenance costs play an important role when considering product demand. Installed price[1] reductions have led to a significant rise in installed systems throughout the past decades (Barbose & Darghouth, 2019). However, how sensitive are residential investors to a change in prices? Which other factors are important when it comes to deciding how much capacity to invest in? And do the answers to these questions vary for different customer groups?

Relevant insights regarding this topic can be of tremendous importance, as they enable federal and state governments to design better tailored, more efficient policy incentives and regulations while producing firms can make

[1] Hereinafter, installed price refers to the total costs for solar PV installation, including hardware costs as well as soft costs (customer acquisition, system design, installation labour, permitting, inspection, etc.) and Balance-of-System costs (racking, wiring, etc.) (Barbose & Darghouth, 2019).

DOI: https://doi.org/10.5282/jums/v7i3pp643-667

informed decisions on product design, pricing, and market forecasts.

1.2. Research Aims

This study's goal is threefold, placing particular emphasis on the first of the three research questions:

(1) How high is the price elasticity of demand[2] for solar photovoltaic systems in the U.S.?
(2) Which other factors impact the installed system size, and in what way?
(3) Can differences be observed for distinct subgroups of the population?

The price for solar PV is probably easier to influence externally than other demand driving factors. Therefore, knowledge of the marginal impact of price changes can be a crucial lever to design better tailored policies and pricing in order to globally increase the demand for and thereby the share of solar energy generation.

To provide sound and data-driven answers to the key questions outlined above, this study focusses on an econometric approach to estimate demand elasticities. The model specification includes instrumental variables as well as regional and time fixed effects. This way, it addresses some of the main challenges in demand estimation, accounting for endogeneity in regressors and heterogeneity on a regional level and over time, respectively (Cui, 2018; Gillingham & Tsvetanov, 2019a). The analysis is mainly based on a subset of U.S. pooled cross-sectional data collected and pre-processed by the Lawrence Berkeley National Laboratory (NBNL) on distributed, grid-connected PV systems for residential as well as non-residential customer segments from 1998 to 2018. The data set contains key attributes of installed systems, including system size, installed price, received financial incentives, location, module technology, and efficiency. A subset of observations is used for the estimation. I supplement these data with information on production factors like polycrystalline silicon prices and installer labour wages, as well as on incentive programs, electricity prices, and income and tax levels[3].

The analysis aims to derive insights that help to provide tangible and actionable policy implications to promote economic investment in solar PV and maximise the benefits of political and commercial interventions. The estimates can further be used to calculate program effectiveness and assess social desirability by comparing the derived costs of carbon emission abatement associated with solar PV rebates and tax incentives to the estimated (social) cost of carbon emissions assumed by the U.S. government (Gillingham & Tsvetanov, 2019a). Also, looking beyond the scope of this study, identifying a valid method of demand estimation might be beneficial to a much wider range of estimation problems, especially for early-stage technologies (Gillingham & Tsvetanov, 2019a).

The remainder of the paper at hand is structured as follows. Section 2 provides a brief introduction to today's U.S. energy infrastructure in general and the prevalence of solar photovoltaics in particular, underlining the importance of the topic in the light of current global environmental challenges. It outlines current deployment and development of solar PV, and gives insights into policies as well as the political environment regarding the technology in the U.S. The following section 3 turns to the estimation of the solar PV demand curve. It sheds light onto insights gained in relatively scarce prevalent literature on the topic and describes the Instrumental Variable (IV) estimation methodology, its application to the problem at hand, relevant data, and the estimation results. Section 4 discusses the political and economic relevance and implications of the findings. Lastly, section 5 critically assesses the study, touches upon limitations, and draws a comprehensive conclusion, including an outlook on potential future research.

2. Solar Photovoltaics in the U.S.

2.1. U.S. Energy Infrastructure and Solar Capacity today

The United States, as the world's largest economy, have a substantial influence on the global energy consumption as well as its consequences. The energy mix in the U.S. is dominated by fossil fuels, with petroleum, coal, and natural gas making up more than 80% of the country's energy production in 2018, and fossil fuel consumption even having increased by 4% relative to previous year levels (EIA, 2019a, 2019b). As a result, the country generates about 15% of the global energy-related CO_2 emissions (Center for Sustainable Systems, 2019). According to the U.S. Environmental Protection Agency (EPA), U.S. GHG emissions, 80% of which are CO_2, have increased by 1.3% since 1990. Most of them stem from burning of fossil fuels in transportation and electricity generation (EPA, 2020).

Notwithstanding, clean energy sources have gained in importance (EIA, 2020a). From the 1990s onward, renewable sources other than hydropower and biomass started to take a share in the U.S. energy mix (EIA, 2020a). Both U.S. production and consumption from non-fossil energy sources reached record levels in 2019, constituting 20% of the states' total energy consumption (EIA, 2020b). A total of 19% of the U.S. electricity was generated from renewable resources that year, thereof 15% by solar power[4], crowding out less efficient or less ecological alternatives such as coal and oil (EIA, 2020a).

[2] The price elasticity of demand is the percentage change in quantity demanded caused by a one percent change in price, moving along the demand curve. The elasticity can be expressed as the slope of the relationship between the natural logarithm of quantity and price.

[3] An overview of selected variables can be found in Appendix A7, Table 18.

[4] One typically distinguishes between two types of solar power: solar thermal and solar photovoltaic (Khan & Arsalan, 2016). Solar thermal converts sunlight into heat which can subsequently be used for multiple purposes. Solar photovoltaic applications, on the other hand, directly generate electricity from sunlight, using a semiconductor technology (Burr, 2014; Singh,

Solar energy is abundant, inexhaustible, and amongst the cleanest sustainable energy source to date (Denholm & Margolis, 2007; Parida, Iniyan, & Goic, 2011), although negative externalities are evidently not absent, arising during fabrication, construction, and operation (Khan & Arsalan, 2016; Nugent & Sovacool, 2014; Raman, 2013).

Solar power is one of the fastest-growing sources of energy, both globally and in the United States. In 2018, solar energy accounted for about 2% of the total U.S. energy consumption, but exhibited a growth of 22% compared to 2017 levels, highlighting the strong focus and large potential of the technology (EIA, 2019c). The country's total installed solar PV capacity has reached over 81 GW[5] in Q1 2020, following extensive investment in the past years (Perea et al., 2020b). For comparison, global PV installations reached 627 GW by the end of 2019 (Feldman & Margolis, 2020). The U.S. installations in 2019 constitutes a 23% year-over-year increase and represent nearly 40% of the total new U.S. electricity generating capacity installed that year (Perea et al., 2020a). Today's capacity was expected to more than double by 2025 before forecasts declined moderately due to the impacts from the coronavirus pandemic, which will most probably cause less utility PV to be built in the coming years (Perea et al., 2020b). However, especially residential solar saw a record-high in capacity additions in 2019, while, on the other hand, non-residential PV growth declined slightly due to unfavourable policy changes in several states (Barbose & Darghouth, 2019; EIA, 2019a).

2.2. Characteristics of Installations

Solar PV installations in the U.S. vary across customer segments[6] in numerous aspects, including system size, efficiency, and module and inverter technology, constitute Barbose and Darghouth (2019) from a representative U.S. data set on PV installations[7]. Overall, systems grew in size, with a median capacity of 6.4 kW for residential and 47 kW for non-residential installations in 2018 (s. Appendix A1, Figure 1). Module efficiency is highest for residential applications, likely due to greater space constraints compared to non-residential sites (s. Appendix A1, Figure 2). The residential share of monocrystalline silicon modules compared to lower-quality polycrystalline silicon is largest (Barbose & Darghouth, 2019), while production nowadays focusses on even more efficient products such as monocrystalline p-type PERC and n-type PERT, also for large-utility-scale systems (Blakers, 2019; Burr, 2014; Platzer, 2012) (s. Appendix A1, Figure 3). Furthermore, efficiency-enhancing module-level power electronics[8] (MLPE) like microinverters or DC power optimizers are used particularly in residential installations as small roof-top systems are constrained most with regard to orientation and flexibility. On the contrary, ground-mounting and tracking[9] are more common for large non-residential and utility-scale installations. According to Barbose and Darghouth (2019), in 2018, only 3% of all residential systems are ground-mounted. Less than 1% can track the sun, even though many residential rooftop installations do not offer the flexibility to freely choose the panel orientation, and only slightly more than half of the systems were oriented southward in 2018. This might be explained by lower rooftop installation costs, higher market penetration, and the fact that systems became economically viable also without an optimal panel orientation or tracking functionality (Barbose & Darghouth, 2019).

2.3. Development of Installed Prices

Overall, renewable energy sources have become the lowest-cost sources of power in many countries (IRENA, 2019). As one of those, solar photovoltaic has made huge steps towards becoming a mature technology throughout the last decade. In the U.S., average PV prices fell by 50% between 2013 and 2018 (IRENA, 2019). This development can primarily be attributed to higher efficiency, lower module prices, and decreasing system costs and is expected to continue in the coming decades (IRENA, 2019). Capital costs of solar PV include hardware costs as well as Balance-of-System (BoS) and soft costs (Barbose & Darghouth, 2019; Elshurafa, Albardi, Bigerna, & Bollino, 2018). While hardware mainly refers to the PV module and inverter, BoS costs include racking and wiring as well as soft costs for customer acquisition, system design, permitting, and labour costs for installation and inspection (Barbose & Darghouth, 2019; Elshurafa et al., 2018). Unlike hardware costs, these BoS costs differ regionally due to their strong dependence on local wage rates, taxes, and competition (Elshurafa et al., 2018; Gillingham et al., 2016).

Barbose and Darghouth (2019) identify several key trends in prices prior to incentives, using U.S. data from 30 states in the past 20 years. In 2019, national median installed prices in the U.S. ranged from \$3.7/W over \$3.0/W to \$2.4/W for residential to small and large non-residential systems, respectively (s. Appendix A1, Figure 4). These persisting differences can mainly be attributed to higher economies

2013). These range between distributed small-scale residential to utility-scale power generation facilities. Solar PV is the more mature and commercially established technology (Khan & Arsalan, 2016).

[5] All energy is expressed in direct current (DC) units. Direct current describes the flow of energy into one direction only. All solar PV nowadays produce DC power (Zainudin & Mekhilef, 2010).

[6] Distributed PV comprises residential as well as non-residential rooftop PV installations of any size and ground-mounted systems of less than 7,000 kW. Non-residential systems are divided into small and large non-residential, with a threshold of 100 kW, in accordance with Barbose and Darghouth (2019).

[7] For a more detailed description, see section 3.4.1.

[8] Direct current power optimizers and microinverters are collectively referred to as module-level power electronics, or MLPE. They have replaced standard string inverters in the past years, 55% of all residential PV systems using some form of MLPE by 2014 (NREL, 2015). Both microinverters and power optimizers can monitor the performance of individual solar panels, rather than the solar panel system as a whole. They improve performance for solar panels by reducing shading losses and the impact of multiple roof planes (Deline, Meydbray, Donovan, & Forrest, 2012).

[9] Tracking is the technical ability of a system to flexibly change its orientation towards the sun compared to fixed-tilt systems (Barbose & Darghouth, 2019).

of scale for larger installations (Barbose & Darghouth, 2019). In the long term, installed prices fell due to the reduction in both hardware, BoS, and soft costs. Modules and inverters made up about 55% of these total cos reductions and fell most between 2008 and 2012 (Barbose & Darghouth, 2019). The remaining 45% can be attributed to reduced BoS costs. While for residential PV, the decrease was mainly driven by hardware cost, non-residential installers benefited to a major extent from reduced BoS and soft costs (Barbose & Darghouth, 2019). The decline in non-hardware costs cannot be linked to a single factor, but rather a changing market and policy environment as well as mechanical aspects. Regarding technical reasons for price reduction, drivers are twofold: On the one hand, installing larger systems on average reduced costs per watt as fixed costs for permitting and customer acquisition occur irrespective of installation size. On the other hand, hardware technologies have improved and – due to the extreme price deflation of PV in the past years – were able to push less efficient polycrystalline modules out of the market. Increased module efficiencies in turn lead to a disproportionately low increase in area-related costs like racking and wage costs for installation of a given capacity (Barbose & Darghouth, 2019). Overall, in 2019, prices continued their declining trend, though at a slower pace. This reduced marginal change is mainly due to lowered financial incentives and higher customer acquisition costs as most early adopters already installed solar PV. Also, cost reductions and efficiency gains become increasingly difficult to capture as the market matures (Barbose & Darghouth, 2019). With rising grid penetration, more significant cost reduction becomes necessary in order to make solar PV profitable for the remaining potential customers, especially in states exhibiting already high PV shares.

Looking at an absolute price level, Barbose and Darghouth (2019) and Gillingham et al. (2016) find that costs vary substantially across states (s. Appendix A2, Figure 9). Smaller markets are generally associated with higher prices but more significant cost reductions. Likewise, urban areas tend to show higher prices. According to Barbose and Darghouth (2019), even on an individual installer-level, median installed prices deviate substantially. They are significantly higher for systems with premium efficiency modules and MLPE as well as ground-mounted systems. Interestingly, tax-exempt customers, mostly non-residential, also exhibit higher prices on average.

Understanding the key drivers and characteristics of PV system price changes as well as distinct regional differences is essential for the following estimation and interpretation of the results, especially because changes in the installed price need to be approximated through variation in correlated (instrumental) variables.

2.4. Incentive Policies and Political Environment

The United States implemented several utility, state, and federal incentive mechanisms to foster growth of solar PV (Barbose & Darghouth, 2019; Consumer Energy Alliance, 2018; Platzer, 2012; Shrimali & Jenner, 2013). Mostly, these comprised of cash incentives through the state or utility PV programs in the form of rebates or grants, performance-based incentives (PBIs), and federal and state investment tax credits (ITC) for both distributed and utility-scale systems. Furthermore, tax exemptions, rights for accelerated depreciation, retail rate net metering, a market for solar renewable energy certificates (SRECs), and non-rebate marketing programs had been established, some of which still exist (Barbose & Darghouth, 2019; Gillingham & Bollinger, 2019; Shrimali & Jenner, 2013). Tax exemptions for schools, governments, and non-profits result in a disproportionately large share of reduced tax costs for non-residential customers, find Barbose and Darghouth (2019). ITC supports investment since 2005 by providing a dollar-for-dollar reduction in tax liabilities (Platzer, 2012; SEIA, 2020). Distributed as well as large-scale utility installations are eligible to tax credits of up to 30% of purchase and installation costs. In 2015, the ITC was extended to 2021 and 2022 for residential and commercial applications, respectively, but the credit value will start to decline in 2020 (SEIA, 2020). However, not only ITC but also cash incentives have decreased throughout the past decade (Barbose & Darghouth, 2019). Many regions plan to phase out local government incentive programs in the coming years or have already done so (Gillingham & Tsvetanov, 2019a). At peak times providing cash incentives of \$4-6/W, those expired in most larger markets or diminished to less than \$0.5/W on average. However, other forms of financial support like SRECs have become more profitable and thus more prevalent (Barbose & Darghouth, 2019). On a state level, Renewable Portfolio Standards (RPS) encourage investment in green technologies by requiring certain contributions of renewable sources to the state's energy generation (Yin & Powers, 2010). SRECs markets facilitate compliance with these obligations. PV system owners have the possibility to sell SRECs from their installations, offering indirect cash incentives. Several states prefer these generation-based incentives over standard offer-based ones (Barbose & Darghouth, 2019). If one can assume that demand side subsidies are directly considered in the purchase decision and supply side subsidies are passed on (at least partly) to consumers (Dong, Wiser, & Rai, 2018; Gillingham & Tsvetanov, 2019a), then the changes in incentives directly impact installed prices and consequently solar PV demand[10].

In spite of numerous programs put in place to promote solar power usage, American energy policy has changed drastically under the current administration. It reversed several former agreements and targets and has, for instance, replaced the Clean Power Plan with a weaker Affordable Clean Energy Rule (Keyes et al., 2019; Krupnick et al., 2018). As the U.S. are one of the largest exporters of crude oil and natural gas (EIA, 2019a), economic interests seem to oppose the

[10] Gillingham and Tsvetanov (2019a) find a pass-through rate of cost reductions from the installer to the consumer of 84%, Dong et al. (2018) even find nearly 100% incentives pass-through for residential customers in California, implying a competitive market and well-operating subsidy programs from a pass-through perspective.

goal of deep decarbonisation and economy-wide emission reductions of 80% by 2050, stated in the U.S. Nationally Determined Contribution (NDC) of the Paris Agreement (Dennis, 2019; United Nations, 2016). On November 4^{th}, 2019, the U.S. government officially announced to withdraw from the Paris Agreement, although their pledge remains legally valid until November 2020 (Dennis, 2019; Zhang, Dai, Lai, & Wang, 2017). As a response, some states formed subnational climate initiatives and continue to strive for the previously set goals (Center for Climate and Energy Solutions, 2019; Friedman, 2019).

Furthermore, the current administration made some changes regarding solar energy in particular. In January 2018, the U.S. government placed a 'Section 201 Solar Tariff' on imported solar cells and modules, rendering investment more expensive – especially for utility-scale applications, as hardware costs increase (SEIA, 2019). These frequent policy and price changes make it an even more pressing matter to understand their effective impact on the demand for more renewable energy sources like solar photovoltaics.

3. Estimation of the Demand Curve

3.1. Evidence on Price Elasticity of Demand

So far, existing research on the demand for solar photovoltaic systems is very limited. Most research rather focusses on price elasticity of electricity demand than on the demand for the generation technology itself (Bernstein & Griffin, 2006; Mewton & Cacho, 2011; Miller & Alberini, 2016). Also, as solar PV can still be considered a maturing technology (Khan & Arsalan, 2016; van der Hulst et al., 2020), demand and supply conditions are constantly changing, making it harder to capture influences that remain valid over time. Recently, Gillingham and Tsvetanov (2019a) were the first to simultaneously address three main empirical challenges in estimating the demand for residential solar PV: price endogeneity, unobserved geographic heterogeneity, and excess zeros in the outcome variable with count data. Using panel data[11] on Census block level from Connecticut on the count of annual solar PV systems installed, Gillingham and Tsvetanov (2019a) account for heterogeneity in block group-specific characteristics by including geographic fixed effects and year dummies (Wooldridge, 2005). Furthermore, the address the issue of excess zeros in count data by applying a two-stage Poisson hurdle model[12] consisting of a logit regression with a control function and a truncated Poisson estimated by a General Method of Moments estimator. They include local roofing contractor wage rates and state incentives for PV systems as instrumental variables to eliminate the endogeneity in the price regressor. Their results suggest that residential consumers are relatively price insensitive (−0.65), meaning that the demand decreases less than proportional to the price increase.

Cui (2018) takes a slightly different approach to estimate both demand and supply functions of rooftop solar panels in California using data from the California Solar Initiatives rebate program. Like Gillingham and Tsvetanov (2019a), she estimates a hurdle model with count data to account for zeros in installation numbers aggregated by zip code and month. Likewise, Cui (2018) assumes a two-part non-linear model and uses a control function instead of a two-stage least squares estimator to account for endogeneity. Employing changes in rebates as exogenous variable to estimate supply and demand function simultaneously, she finds very different results compared to Gillingham and Tsvetanov (2019a), with a demand elasticity of -3.824 and supply elasticity of 5.572. She also specifies one model estimating the system size instead of installation count but does not obtain significant estimates. According to Cui (2018), customers are highly price sensitive, wherefore rebates are a very effective way to promote PV adoption. Cui (2018) further states that elasticity is not constant, but that consumers and sellers get more price inelastic as prices decrease. Besides the different model specifications, regional characteristics might to some degree provide explanations for the deviating results of these two studies. Gillingham and Tsvetanov (2019b) state that, unlike in Connecticut, the phase-out of subsidies could be anticipated beforehand in California, probably impacting the timing of investment decisions. Another aspect leading to differing estimates could be the limited sample of rebate installations as well as the slightly less granular assessment on zip code rather than Census block level.

Exploiting the changes in rebate rates for residential systems in California, Hughes and Podolefsky (2015) use a reduced form equation to estimate the number of installations. They find relatively high rebate elasticities of about -1.2, accounting for mean and utility specific unobservable characteristics that affect PV adoption and vary over time. Rogers and Sexton (2014) conclude that rebate elasticities are slightly lower, estimating a reduced form rebate elasticity of -0.4 for California.

Overall, the need for further research becomes apparent, as prevalent insights are both divergent and scarce. Numerous research designs are employed, ranging from different predictors over various forms of model specifications and estimation methodologies. Therefore, there is no consensus on typical demand curve characteristics and elasticities for solar photovoltaics so far. To my knowledge, there is no study yet providing substantial insight on the price elasticity of solar photovoltaic installations' system size installed, as will be the focus of this study.

3.2. Methodology and Research Design

3.2.1. Issues in demand estimation

Estimating demand and supply curves and their factor elasticities inherently poses the issue of simultaneous causal-

[11] Panel data are multidimensional data that include measurements pooled over space and time. They are a combination of cross-section and time series data (Baltagi, 2008).

[12] Hurdle models are motivated by sequential decision-making. The can represent the process of first deciding whether to buy or not and secondly deciding on the (positive) quantity to buy (Gillingham & Tsvetanov, 2019a).

ity[13] because the observed data of prices and quantities represent a set of market equilibria where supply equals demand (Angrist & Krueger, 2001). The price of a good influences its quantity bought and vice versa. An ordinary least squares (OLS) regression is incapable of isolating the effect of a price increase on one of the two curves, making alternative methods of estimation indispensable (Stock & Watson, 2020). A general linear model of solar PV demand would be specified as follows (Eq. (1) and (2)):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i \tag{1}$$

Or, in matrix algebra

$$\boldsymbol{Y} = \boldsymbol{\beta X} + \boldsymbol{u} \tag{2}$$

For $i = 1, \ldots, n$ observations, let Y denote the dependent variable to be estimated: the size measured in watts of a solar PV system installed. X represents the k demand shifters or regressors incorporated in the model to estimate Y. The βs measure the effect size of the respective variables and are the coefficients of interest. The demand function is assumed to be linear in its parameters here.

A standard OLS approach to estimate the demand curve (Eq. (1)) makes several assumptions. One of them is that regressors $X_{1i}, \ldots, X_{ki}$ and error term u_i are uncorrelated, i.e., the conditional expectation of the error given the regressors is zero ($E[u_i | X_i] = 0$). Those regressors are called exogenous. In other words, there is no unobserved variable that is correlated with X and simultaneously changes Y. If this does not hold true, resulting OLS estimates are inconsistent even for large samples and other methods of estimation are necessary (Stock & Watson, 2020). This is where one challenge arises in this study. As for every supply-demand problem, the above-stated assumption of exogenous regressors is violated. An OLS estimation is biased because the simultaneous causality induces a correlation between the price regressor and the unobservable error term. Thus, how can this demand function be estimated consistently with the present endogeneity in prices? First, I define the model more precisely, separating endogenous and exogenous regressors:

$$\boldsymbol{Y} = \boldsymbol{\beta X} + \boldsymbol{\gamma W} + \boldsymbol{u} \tag{3}$$

Y still denotes the system size. The right-hand side combines its determining factors as well as the error term u. X now represents the endogenous regressors correlated with u while W denotes the truly exogenous regressors which are not correlated with u. In this estimation, only price per watt is assumed to be endogenous. With an endogenous regressor X the estimate of β will be incorrect (Stock & Watson, 2020). It captures both the effect of independent changes in X as well as changes in the error u due to simultaneous changes in Y associated with X.

[13] Simultaneous causality means that "causality runs 'backward' from Y to X as well as 'forward' from X to Y" (Stock & Watson, 2020, p. 428). It is one cause for endogeneity in regression problems.

There are several possible solutions to obtain consistent estimates in the case of simultaneous equations (Wooldridge, 2015). One, and probably the most frequently used one, is an Instrumental Variable approach, first employed by Philip Wright in 1928 (Angrist & Pischke, 2008; Wright, 1928).

3.2.2. Theory of instrumental variable estimation

The basic idea behind an IV estimation is to eliminate any correlation of the endogenous regressors X with the error term u by finding other variables that can be used in the regression in their stead. These variables are called instruments. Demand and supply estimation problems were the first applications of instrumental variables, initially called 'curve shifters' (Angrist & Krueger, 2001). For demand estimation, these curve shifters are used to trace out the slope of the curve by an exogenous variation in the supply, modifying cost conditions without affecting demand conditions (Angrist & Pischke, 2008; Stock & Watson, 2020).

An instrument, let it be Z, needs to satisfy two conditions of validity in order to produce meaningful results (Stock & Watson, 2020): Firstly, it must be *relevant*, i.e., highly correlated with the endogenous regressor to be replaced in order to serve as a good proxy variable ($Cor(Z, X) \neq 0$). The more variation in X can be explained by variation in the instrument, the more information will be sustained in the IV regression. If an instrument explains only a minor part, it is called weak (Stock & Watson, 2020). In this case, a weak instrument could be the price for an input factor that accounts only for a very small share of the final price or – with heterogeneous outcomes – influences only few observations, like the price for a rare module technology. Secondly and equally important, an instrument must be *exogenous*. This means that it must not affect the left-hand side of the equation – the variable to be estimated ($Cor(Z, u) = 0$). The variation in X that is related to Z is not related to u, neither through a direct effect of Z on Y nor through a variable that is not included in the model but is causally linked to both Y and Z (Stock & Watson, 2020). Given these assumptions hold true, the exogenous part of the variation in the endogenous regressor X can be isolated via changes in Z and can subsequently be used for a consistent estimation of Y (Angrist & Pischke, 2008; Stock & Watson, 2020). For this to work, one endogenous regressor must have at least one but can have several instruments. An IV model is called overidentified if the number of instruments exceeds the number of endogenous variables (Stock & Watson, 2020). Otherwise, it is exactly identified. Both relevance and exogeneity of instruments can be tested statistically, at least for overidentified models, e.g., by using the first stage F-statistic and the test on overidentifying restrictions (J-Statistic), respectively (Stock & Watson, 2020).

According to Angrist and Krueger (2001), the most efficient way to obtain estimates in an IV regression, especially when using numerous instruments, is the two-stage least squares (TSLS) approach. It builds upon two OLS regressions run in a row to compute the TSLS estimators.

(1) In the first stage, the part in the variation in X that is uncorrelated with the error u is isolated by regressing the

endogenous variable X on the instruments Z and all further exogenous variables W. X is split into two components: one that is correlated with the error and one that is not. For each endogenous regressor – in this study only the price –the reduced form equation (4) needs to be estimated by OLS (Stock & Watson, 2020):

$$\boldsymbol{X}=\pi\boldsymbol{Z}+\delta\boldsymbol{W}+\boldsymbol{\nu} \tag{4}$$

(2) In the second stage of TSLS, the idea is to estimate fitted values $\hat{X}$ for all observations building on the first stage results and subsequently use these fitted values $\hat{X}$ instead of the original values X for estimation of the actual model specification. The original equation (3) is estimated again by OLS, only that X is replaced by $\hat{X}$:

$$\boldsymbol{Y}=\boldsymbol{\beta}\widehat{\boldsymbol{X}}+\boldsymbol{\gamma}\boldsymbol{W}+\widetilde{\boldsymbol{u}} \tag{5}$$

This regression provides consistent estimates $\widehat{\beta}^{TSLS}$, converging asymptotically towards the true parameter as the sample size increases (Stock & Watson, 2020). Consequently, researchers should work with significant sample sizes when applying an IV approach. However, if the explanatory variable is in fact not endogenous, both TSLS and OLS estimator are consistent, but the latter is more efficient. Therefore, it is important to ensure the presence of endogeneity, for instance by using the Durbin-Wu-Hausman test (Baum, Schaffer, & Stillman, 2007).

3.2.3. Potential issues in instrumental variable estimation

Like many other statistical models, an instrumental variable approach makes some model assumptions that need to be satisfied in order to obtain valid estimates (Stock & Watson, 2020). For IV models, these assumptions are modifications of the OLS assumptions for causal inference outlined in Stock and Watson (2020). In practice it is often very hard to meet all the requirements. Therefore, their validity in this application is discussed in Appendix B1.

More generally, Angrist and Krueger (2001) point out that IVs can solve the first-order problem to remove omitted variable bias[14] (OVB) only for a well-defined population. With heterogeneous responses, not every single observation can be explained by variation in the instruments as they only capture part of the true variation in prices. For example, where module costs do not differ significantly but mainly other hardware, BoS, or soft costs are drivers of price changes, an instrument shifting module prices will not accurately depict the price variation. Including several instruments can possibly counteract this to some extent. However, due to the bias-variance trade-off (James, Witten, Hastie, & Tibshirani, 2013), using more instruments might increase the variance of the estimators (Angrist & Krueger, 2001).

Lastly, another pitfall in IV estimation mentioned by Angrist and Krueger (2001) are functional form issues for both stage estimations. They emphasise that in a TSLS estimation procedure, the consistency of the final estimates $\widehat{\beta}^{TSLS}$ does not depend on the correct functional form of the first stage regression (Kelejian, 1971). Therefore, I estimate a linear regression for the first stage as a more complex non-linear model does not generate consistent estimates unless the fit is exactly right (Angrist & Krueger, 2001).

3.3. Application to Solar PV Demand Estimation

3.3.1. Specifying a multiple log-log linear regression model

The difficulties of demand estimation problems as well as one possible solution to solve them have been introduced in the previous sections. In the following, I apply this to the estimation of solar PV price elasticity. The preferred model[15] specification in this study is a multiple linear regression with log-transformed continuous outcome and predictor variables, including time and regional fixed effects on state and year level, controlling for several potentially confounding variables, and instrumenting for price. It is specified as follows:

$$\log(\boldsymbol{Y})=\boldsymbol{\beta}\log(\boldsymbol{X})+\boldsymbol{\gamma}\log(\boldsymbol{W})+\boldsymbol{\alpha}+\boldsymbol{\mu}+\boldsymbol{u} \tag{6}$$

Y is the system size installed in watts, X represents all explanatory variables (exogenous and endogenous), W are control variables, and α and μ are fixed effects. The primary objective is to assess the constant elasticity of the system size installed with respect to the explanatory variables. Therefore, I estimate a log-log additive linear model (Eq. (6)). Its coefficients can directly be interpreted as an expected percentage change in system size given a regressor increases by one percent (Benoit, 2011). Rather than focussing on absolute differences, I estimate the relative change, the elasticity. If the price elasticity is constant, a percentage increase in price can be expected to cause a proportionate change in PV size demand over a wide range of prices. Thus, assuming a constant elasticity implies that absolute changes in system size may differ depending on the former level. This seems reasonable because a price increase by 1% when prices are low might well induce a shift in system size installed that is different from the shift caused when initial prices are relatively high. This means that a log transformation natively handles non-linear relationships between system size and independent variables (Benoit, 2011; James et al., 2013). For the data used here, regression plots show that the relationship between system size and the explanatory variables after log-transformation is at least slightly more linear (s. Appendix A4, Figure 18 to Figure 21). Apart from simplifying the functional form, log-transformed variables often follow an approximately normal distribution for otherwise skewed variables. In this case, several variables are right skewed

[14] "If the regressor [...] is correlated with a variable that has been omitted from the analysis [...] and that determines, in part, the dependent variable [...], then the OLS estimator will have omitted variable bias. Omitted variable bias occurs when two conditions are true: (1) the omitted variable is correlated with the included regressor and (2) the omitted variable is a determinant of the dependent variable." (p. 212 Stock & Watson, 2020).

[15] Hereinafter, preferred and final model synonymously refer to the model selected as the best model after conducting diagnostic and model validity tests as well as plausibility considerations.

(s. Appendix A4, Figure 22 and Figure 23), wherefore a transformation seems appropriate in order to better satisfy the assumptions of model linearity and normal distribution of the errors.

The linear additive model of the logs holds further advantages over other specifications. It can easily incorporate fixed effects to account for heterogeneity and instrumental variables to eliminate endogeneity (Gillingham & Tsvetanov, 2019a). Furthermore, its coefficients allow a straightforward interpretation, making it very popular with many researchers.

3.3.2. Deciding on the dependent variable

As described in section 3.1 above, most former research estimates demand models with installation count instead of system size as dependent variable (e.g., Cui, 2018; Gillingham & Tsvetanov, 2019a; Hughes & Podolefsky, 2015). Rather than determining to what extent price changes influence the system size installed they assess the propensity of adoption, meaning the decision to invest in solar PV, measured by the number of installed systems in each area. Unlike this study, those consequently also consider a zero-realisation with the decision not to invest by using a two-stage hurdle model. Thereby, they account for the fact that changes in demand factors can incite markets to change from zero to nonzero quantities, and vice versa. However, in order to use count data, a very broad and high-quality data coverage is necessary such that a count of zero installations can be attributed to the decision not to buy rather than missing data. Additionally, in California, subsidies were phased out in a way that depended on the total amount of installed PV capacity, allowing consumers and firms to reasonably anticipate the timing of subsidy declines and leading to bundled installations shortly beforehand which would have to be accounted for when using count data (Gillingham & Tsvetanov, 2019b).

Other research applies a dynamic discrete choice[16] approach (e.g., Bollinger & Gillingham, 2019; Burr, 2014), which might seem reasonable since most people would only install a system once. However, Gillingham and Tsvetanov (2019b) provide evidence that the investment in solar PV can more often be treated like normal purchases rather than a "buy-or-wait" decision. Furthermore, as more capacity can be added later, the decision on the size of the solar PV system is not necessarily a discrete choice. For those reasons, this study takes system size as the dependent variable to be explained.

3.3.3. Identifying suitable instrumental variables

Knowing of the endogeneity issue in the estimation, how can the IV approach be applied to consistently estimate the model elasticities β? The first step is to find suitable instruments for the installed price per watt. Generally, the choice of good instruments first and foremost relies on a profound understanding of the economic mechanism behind the relationship of interest (Angrist & Krueger, 2001; Angrist & Pischke, 2008). This leads to plausible and more intuitive results compared to abstract theoretical models based on hard-to-verify assumptions about certain distributions and relationships, so Angrist and Krueger (2001). To trace out the demand curve, one or more instruments that impact the supply of PV modules but not their demand have to be found. A typical supply-side instrument shifts costs of sales or production as these can be expected to affect supply without impacting demand (Angrist & Krueger, 2001; MacKay & Miller, 2019). Although statistical tests can assist to evaluate the relevance and exogeneity of instruments, it is useful to think about whether a chosen instrument plausibly satisfies these conditions (Stock & Watson, 2020). In complex demand models, especially the exogeneity of instruments can be challenging to assess. In this study, the preferred model employs two instrumental variables to approximate changes in the price for solar PV: Polycrystalline silicon prices and the sales tax rate on solar PV installations.

The first instrument, price quotes for polycrystalline silicon or 'polysilicon', aims to capture the variation in input factor costs. Polysilicon is the main raw material used for PV module production (Woodhouse, Smith, Ramdas, & Margolis, 2019). According to Woodhouse et al. (2019), crystalline-silicon made up about 90% of all PV production in 2014, increasing to over 95% in the subsequent years. Polysilicon is the basis for production of both multi- and monocrystalline silicon ingots of different purity and efficiency levels, which are then processed to wafers, manufactured into cells, and eventually fabricated into entire PV modules. Therefore, nearly all PV modules installed in the market in the past years used polysilicon as one factor of production. Moreover, the installed price of solar PV is to about 55% determined by hardware component costs (e.g., module, inverter) and to 45% by BoS and soft costs (e.g., installation labour, acquisition cost, system design, permit and inspection, installer margins, loan-related fees). As the following estimation considers only residential systems, for which the price decline in the past years was mainly driven by hardware cost reduction, instrumenting price changes via variation in those costs seems reasonable (Barbose & Darghouth, 2019). Therefore, input factor prices can be assumed to represent a *relevant* instrument. Further, as it seems plausible to assume that a residential customer's demand does not directly rely on the level of the price for polysilicon, these should be *exogenous* and influence demand only indirectly through the installed price. Even though there are possibly several relevant variables omitted from the model, these are similarly unlikely to be related to global polysilicon prices and will, therefore, not cause a correlation of this instrument with the error. Notwithstanding, a potential limitation of the power of the polysilicon prices as an instrument could be the fact that these are relevant only at the very beginning of a complex and costly production chain, making up less than 10% of the production costs of monocrystalline PV modules in 2018 (Woodhouse et

[16] A dynamic discrete choice model estimates the decision of a forward-looking agent over a finite number of options who is taking the utility of future alternatives into account (Heckman & Navarro, 2007).

al., 2019). Consequently, final module prices could differ significantly reflecting costs of subsequent production steps or further input factors even for initially equal polysilicon input costs, leading to a relatively weak instrument. This would imply that further production cost factors for PV modules need to be incorporated as instruments in the first stage estimation to approximate the price development more precisely. However, further data on hardware production costs are limited and soft costs are rather hard to quantify as they differ severely, depending on changing market and policy environments (Barbose & Darghouth, 2019). To account for a part of cost fluctuations for different modules and inverters, I include dummies indicating the type of module technology as well as MLPE in the model, where this information is available.

The second variable to instrumentalise installed prices is the sales tax rate levied on the hardware costs of installed systems, again assumed to be 55% of the total installed price. Tax rates are a popular instrument for price changes (Frondel & Vance, 2013; Stock & Watson, 2020). For one thing, they can be assumed to be *relevant*, as after-tax sales prices are adjusted to incorporate changes in taxes and often make up quite a noticeable part of the price for the final customer. We would assume prices to increase alongside rising sales tax rates, expecting producers to pass on at least part of the additional costs to consumers. The question whether tax rates can be considered *exogenous* is somewhat harder to answer. On the one hand, a change in tax rates can plausibly be assumed to impact consumer demand solely through the adjustment in prices and not directly through the mere fact that tax rates changed. However, unlike polysilicon prices, tax rates are more likely to be related to some of the omitted variables in the error term. While unobserved variables like hours of sunshine and irradiance are unrelated to setting tax rates, electricity prices may change alongside tax rates and simultaneously influence the amount of solar PV invested in. Also, incentive payments could be linked to tax policy, because general sales tax as well as many incentive programs are determined by state governments through political and financial considerations (Tax Policy Center, 2020). If these are not entirely captured by fixed effects included in the model, this might lead to inconsistent TSLS estimates. For this reason, the exogeneity test result will be particularly relevant here. In this study, the sales tax rate includes any potential tax exemptions granted to solar PV investors through the state government (Barbose & Darghouth, 2019; Shrimali & Jenner, 2013) which might lead to other-than-expected coefficients in the first stage linking tax rate and price per watt, if tax exemptions are primarily granted where prices are higher.

Apart from the instruments described in more detail above, I estimate specifications using further potential supply shifters given in section 3.4.2.2. However, mostly due to missing data and limited granularity, these did not yield meaningful estimates.

3.3.4. Defining relevant explanatory variables and controls

Other than price, further aspects determine the size of a solar photovoltaic system to be installed. In this estimation, I include information on the module efficiency assessing the energy conversion efficiency of the modules, and dummies indicating ground-mounting, tracking functionality, and the fact whether the system is installed as retrofit on an existing house or during the construction of a new building. Apart from these, data on the kind of module technology and MLPE are incorporated in the model. Module technologies are grouped into polysilicon, monocrystalline silicon, or other technologies. As MLPE categories I consider microinverters, DC optimizers, or no power electronics.

Naturally, there might be a large range of other variables that could turn out relevant determinants of the system size installed. However, as for the decision to invest in a good, especially a more complex one like a solar PV system, investors take many different variables into account, including the data on all these will probably be unattainable. This is no major cause for concern if the omitted factor is not correlated with any variable included in the model. However, if it is, this variable's coefficient estimate will be biased, reflecting not only its own effect but also that of the omitted variable (Stock & Watson, 2020). For this reason, control variables, though their coefficients are not of primary interest and might not have a sound causal interpretation, need to be incorporated in a model if their absence would otherwise cause OVB (Angrist & Pischke, 2008; Stock & Watson, 2020). In the final model, I include information on the number of households and the adjusted gross income per household on zip code level per year. A high level of income might positively relate to the propensity to invest in solar PV and its system size while the number of households, approximating the population and building density, could negatively impact the system size installed. In addition, they could in numerous ways be related to predictive variables included in the model, e.g., the efficiency, tracking equipment, or ground-mounting, assuming that richer people can afford to buy higher-class modules and higher population density requires the purchase of roof-mounted installations. By including the information in the model, I avoid confounding effects in the estimates of my coefficients of interest (Stock & Watson, 2020).

3.3.5. Accounting for heterogeneity through fixed effects

In a pooled data setting, fixed effects allow to eliminate OVB caused by factors that are not included in the model and which vary across states, but are constant over time (state fixed effects α_i), or which vary over time, but are constant across states (time fixed effects μ_t) (Borenstein, Hedges, Higgins, & Rothstein, 2010; Stock & Watson, 2020). They measure the residual difference across state and time, respectively, after accounting for all other factors in the model. Including an interaction term between state and time fixed effects allows the time effect to be different for individual states and vice versa (Stock & Watson, 2020). In this study, state fixed effects might capture e.g., weather conditions and

hours of sunshine, established institutions, as well as prevalent fundamental culture and values. Time fixed effects, on the other hand, can account for aggregate time-varying demand shocks across states, U.S. economic and population growth, inflation, technological progress, and federal policy changes. One needs to bear in mind that time fixed effects might capture the impact of varying production factor input prices as well, if those are not included in the model. The interaction of both state and time fixed effects represents how state-specific factors change over time, e.g., state legislature and regulations, and subsidy and tax policies, as well as market conditions, prominent mindsets, trends, and acceptance of innovative technologies in a state. These interaction effects will only change estimations if enough data from different states and time frames is included, however. After data selection, this does not hold true for this study. The final model is estimated without interaction as there was no significant difference in the coefficient estimates.

3.4. Data

3.4.1. 'Tracking the Sun' data

Data set and structure

The original data of which a subset is used for estimation were collected and pre-processed in the 'Tracking the Sun' (TTS) data set by the Lawrence Berkeley National Laboratory (LBNL). Overall, the data cover PV systems installed in the U.S. from 1998 throughout 2018 with trends of the first half of 2019 (Barbose & Darghouth, 2019) (s. Appendix A2, Figure 5). The installation data were primarily reported to state agencies and utilities managing PV incentive programs, SREC registration, or interconnection processes in 30 states. The sample contains project-level information on nominal installed prices, system size, tax payment, financial incentives, module and inverter technology, efficiency, location, and further relevant characteristics of grid-connected, distributed solar PV systems. It excludes utility-scale systems. Barbose and Darghouth (2019) also dropped duplicate observations[17] and those where information on system size or installation date were missing. They corrected the data for obvious errors and standardised installer, module and inverter labels. Overall, 1,543,831 PV systems are included in the full sample, making up about 80% of all U.S. distributed PV systems installed throughout 2018 (Barbose & Darghouth, 2019). Most data stem from California, as solar PV is most prevalent here. In line with this, sample coverage tends to be weaker in small and mid-sized state markets.

Data selection

To improve accuracy, interpretability, and generalisability of the demand estimation, only a subset of the full TTS data sample[18] is used. First, I drop all observations with invalid price information. For this purpose, I exclude systems where installed prices are missing, which are about 23% of all observations. Among the remaining, I filter extreme outliers by including only installations with an installed price per watt between 1 and 10 USD. Those are prices between the 1^{st} and 99^{th} percentile in this data set, rounded inward to the nearest integer (s. Appendix A3, Figure 10). Likewise, all third-party owned systems (39.4% of full sample), systems with appraised price values (25.4% of full sample), self-installed systems (1.4% of full sample), and systems with battery backup (0.7% of full sample) are not considered in the estimation data set. All of these generally exhibit less representative price quotes (Barbose & Darghouth, 2019; Gillingham & Tsvetanov, 2019a). By excluding them, I prevent outliers and erroneous or unrepresentative values to distort the estimates. Thereafter, 613,157 observations (about 40% of full sample) remain in this price sample.

Apart from invalid price observations, further installations are removed from the estimation data set. To start with, I focus on residential systems only, because as shown above, installed prices and installation characteristics vary substantially across customer segments (Barbose & Darghouth, 2019) and residential systems make up almost 95% of the full sample and 93% of the price sample, respectively (s. Appendix A2, Figure 6). Also, price variations in residential PV systems are more likely to be captured by the instruments as their price decline was mainly driven by hardware cost reduction, whereas non-residential installers benefited to a major extent from lower BoS and soft costs (Barbose & Darghouth, 2019). Again, to account for outliers and in line with the customer categorisation by Barbose and Darghouth (2019), installations with system size bigger than 20,000 W are excluded (s. Appendix A3, Figure 11). As I assume efficiency to be non-negative, I also drop all observations with an efficiency of less than 0%. Those are less than 0.1% of the full sample and look like reporting errors rather than true values. Lastly, I consider only systems that were installed after 2009 (94% of full sample, 89% of price sample), as prior to this, data coverage in the sample is very limited and complementary instrument data on polysilicon prices is not available on a weekly basis.

All in all, these selection criteria lead to the estimation sample with 501,394 observations from 21 states and 9 years, representing 32.5% of the full TTS sample. Before being able to use this data for estimation, several data pre-processing steps need to be taken – namely the transformation and generation of predictors. Please refer to the Appendix B2 for a description of the most important steps.

[17] Few duplicate systems with redundant information were left in the sample. Those were deleted to ensure that one installation from one point in time is only included in the sample once.

[18] Hereinafter, full sample refers to the 'Tracking the Sun' data set as published by the LBNL (Barbose & Darghouth, 2019). Price sample refers to the sample left after applying all price-related selection criteria described. Estimation sample refers to the sample left after applying all selection criteria, also non-price related. The final sample is the sample left for model estimation after dropping all observations which have missing values in one or more of the included variables described in section 3.5.1.

Apart from the *Tracking the Sun* data set, further instrument as well as control data are included in the estimation. I elaborate on the more relevant in the following.

3.4.2. Complementary data

Polysilicon price data

Besides the tax rate, which is extracted directly from the TTS data set, I use the movement in price quotes for polysilicon as means to instrumentalise price changes in solar PV. Therefore, these need to be added to the estimation sample. There are different indices and data providers for polysilicon spot prices. Polysilicon production is concentrated almost entirely in China (Platzer, 2012; Woodhouse et al., 2019). Hence, figures referring to China commodity prices can be applied globally. I compare four different data sources providing weekly price quotes extracted from Bloomberg for China, international spot outside mainland China, companies regardless of region, and PVinsights poll prices for multiple contributors[19]. They are provided in USD per kg for comparable polysilicon purity from end of 2009 to mid-2019. However, only the PVinsights data are complete for the years 2010-2018. As those are correlated at over 99% with polysilicon spot prices from the other three indices and all but the data for China exhibit very similar median prices (s. Appendix A3, Figure 12 to Figure 14), I only use PVinsights poll prices in the subsequent estimations.

Despite the fact that in the estimation sample, polysilicon modules make up only about 35% of all observations, compared to 38% monocrystalline silicon, the price movement can be expected to be similar, as polysilicon is the fundamental input material for both module types (EIA, 2019d; Woodhouse et al., 2019).

Further instrument data

Although in the final estimation, only polysilicon price movement and changes in the sales tax rate are used to estimate installed PV price movements, I evaluated the quality of other instrument data, which I only outline briefly.

As next to hardware cost, BoS and soft costs for e.g., assembly, installation, and wiring make up about 45% of installed prices, PV costs can be expected to move in line with solar PV installer wages. Furthermore, changes are probably exogenous as those wages neither influence demand directly nor through other relevant variables, once accounting for income. Therefore, I consider the mean hourly and annual wage rates for U.S. solar PV installers estimated in the Occupational Employment Statistics Survey by the U.S. Bureau of Labor Statistics (BLS) for the years 2012 to 2018 in the estimation. Unfortunately, these data were not available on a more granular level regarding time frame and location, and did, therefore, not capture enough variation to serve as decent instruments for installed prices.

Another instrument I perceived promising is the number and scope of incentive programs, taken from the Database of State Incentives for Renewables and Efficiency (DSIRE). Because financial compensation reduces the effective installed price, it can be expected to provide a *relevant* instrument. However, those programs might reduce the marginal costs by also affecting other demand or supply conditions (Gillingham & Tsvetanov, 2019a), e.g., if there were a link between lower sales tax and offering incentives to promote consumption. Moreover, as the data available were highly incomplete with respect to start and end date as well as the size of financial benefit, I could only include the overall number of programs per state, which is by far not detailed enough to provide an accurate and valid approximation of the PV price changes.

Apart from these additional external data sources I also assess the absolute amount of sales tax paid per watt as well as the rebate or grant provided per watt contained in the TTS data set as instruments. I discard the former because it shows a substantially lower coefficient of determination[20] R^2 in the first stage compared to the tax rate and might additionally suffer from endogeneity due to the measurement in absolute dollar values. The latter could not be used because it did not provide enough data points for sufficient variation, with almost 10% missing values and over 50% zero rebate or grant. For those reasons, those instruments were not used in the final model.

Control variables

It is almost impossible to include all relevant predictors in a model, be it for the complexity and number of variables or for the unavailability of the required data. Nevertheless, some relevant information can be obtained from publicly available sources and included in the data in order to improve the estimation and prevent bias through omitted variables. An overview over the variables added to the estimation sample can be found in Appendix A3, Table 3. As outlined in section 3.3.4, I use yearly income and personal tax data gathered by the U.S. Internal Revenue Service (IRS) on zip code level for the years 2008 to 2017. More specifically, I join information on adjusted gross income, wages and salaries, the number of households approximated by the number of returns, the population approximated by the number of personal exemptions, the taxable income, and the income tax paid. As many of these variables showed an almost perfect correlation, I keep only the number of households and the adjusted gross income as controls. The data is available on a yearly basis for the whole time period considered in the estimation.

[19] Bloomberg indices: SSPSPSNC (BNEF survey), SSPSPSNI (BNEF survey), SSPFPSNO (BNEF survey), SOLRAPS (PVinsights poll).

[20] The fraction of sample variance of the dependent variable that is explained by the model, i.e., the variance in the regressors (Stock & Watson, 2020).

Moreover, I add the average electricity price for end customers by state and year provided by the U.S. Energy Information Administration (EIA). However, this did not add substantial quality to the estimation as electricity prices differ regionally on a smaller-than-state level and vary significantly throughout the year. Therefore, the data did not accurately picture the relevant electricity price movements and did not significantly impact the demanded system size which is why I do not include this variable in the final estimation.

Some data transformations are conducted to use the sample for the estimation of the log-log linear model. For detailed information, please refer to Appendix B3.

3.5. Estimation Results

3.5.1. Preferred econometric model

The model specification selected for the final estimation of the price elasticity of demand is given in equation (7). As already outlined above, it comprises several technology-related factors as well as economic control variables and fixed effects for state and installation year. The installed price is instrumentalised by the polysilicon spot price and the sales tax rate.

$$\begin{aligned}\log(\text{systemsize}) = {} & \beta_1 \log(\text{price}) + \beta_2 \log(efficiency) \\ & + \beta_3 D_{\text{new}} + \beta_4 D_{\text{tracking}} + \beta_5 D_{\text{ground}} \\ & + \beta_6 D_{\text{technology}} + \beta_7 D_{\text{mlpe}} \\ & + \gamma_1 \log(\text{households}) + \gamma_2 \log(\text{income}) \\ & + \alpha + \mu + u\end{aligned} \tag{7}$$

To estimate this model, 172,106 complete observations are used (11% of full TTS sample). It must be noted that these observations no longer represent the entire U.S. because by far not every state included in the initial sample has valid observations for all relevant variables. Thus, the data used for estimating the model is representative at most for the states included, namely California, Texas, Arizona, and other states initially grouped as there were not enough observations available at the individual state level. Since a major part (99%) of the remaining sample stems from California, the results mostly picture the situation present there rather than the whole U.S. Likewise, only the years 2010 to 2017 are still represented in the data and only a minor share of installations uses tracking or is ground-mounted. Summary statistics of the remaining data are provided in Appendix A5. The estimation results, models with different instruments and subsets of predictors, a comparison to alternative functional specifications to check the robustness of the results, as well as estimations for regional and economic subgroups is provided in the following to answer the three research questions introduced above.

3.5.2. Price elasticity of demand

The final model yields a constant long-term price elasticity of residential demand for solar PV of -0.443 (Table 1, model (3)). A 1% increase in the price per watt reduces the system size by about 0.44%. This means that investors are rather insensitive to price changes, as the system size declines less than proportional to the increase in price per watt. The finding is broadly in line with the slightly higher elasticity estimate of -0.65 obtained by Gillingham and Tsvetanov (2019a) for Connecticut and slightly lower elasticity of -0.4 found by Rogers and Sexton (2014) for California. However, this comparison of estimates is not entirely valid for two main reasons. Firstly, both Gillingham and Tsvetanov (2019a) and Rogers and Sexton (2014) estimate the effect of a price increase on the adoption, i.e., the number of installations, rather than the system size. Among those who purchase a solar PV system, the price elasticity of system size demand could be lower since the decision on how much capacity to buy might be less sensitive to changes in the price level than the decision to buy at all or not. Secondly, the final data set contains mainly Californian installations where price elasticities could be lower as found by Rogers and Sexton (2014). Thus, the estimate is not directly comparable to Connecticut.

3.5.3. Impact of further relevant variables

Explanatory variables

Next to prices, further variables turn out to be highly relevant in determining the system size to be purchased, all parameters being significant at a 0.1% level in the preferred model (3). Generally, for the interpretation of regression coefficients two aspects need to be considered: the significance of an effect and the effect size itself. Findings might be significant in terms of p-value, but this does not imply a practical significance in terms of effect size. Large data sets like the estimation data used here tend to produce highly significant estimates already for very minor effect sizes as very small differences can be detected as sample size rises (Lin, Lucas Jr, & Shmueli, 2013).

Increasing *module efficiency* by 1% results in a 0.41% increase in the installed system size. In absolute terms, this percentage change is almost as large as for installed price per watt. At first glance, the direction of the effect might seem somewhat counterintuitive: with increasing efficiency, less solar PV capacity should be sufficient to generate a given amount of energy, implying that, where there is little space and smaller systems are installed, higher efficiency modules are purchased, and vice versa. On the other hand, however, higher efficiency also makes the whole installation more profitable and worthwhile investing in, possibly inducing investors to purchase more and larger models. Furthermore, higher module efficiency results in relatively lower BoS and soft costs for installation per generated kWh of electrical energy. As these are not explicitly modelled here, the efficiency coefficient might incorporate this positive effect.

Strongly related to the module efficiency are both the type of *module technology* and *module-level power electronics*. As these factor variables are not log-transformed they cannot be interpreted as elasticities but rather as an expected

Table 1: IV regression results for the final model specification and different robustness checks

	Instrumental Variable Estimation Results					
	Dependent Variable: System Size (W)					
	IV: Polysallicon Price	IV: Polysilicoa Price, Tax Rate	IV: Polysilicon Price, Tax Rate	IV: Polysilicon Price, Tax Rate, Rebate/Grant per Watt	IV: Polysilicon Price, Tax Rate, Rebate/Grant per Watt	IV: Polysalicon Price, Tax Rate, Rebate/Grant per Watt Tax per Watt, Hourly Mean Installer Wage Incentive Programs
	(1)	(2)	(3)	(4)	(5)	(6)
Price per Watt	−0.600***	−0.452***	−0,443***	−0,447***	−0.454***	−0.431***
	(0.165)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Module Efficiency	0.175***	0.212***	0.408***	0.213***	0.212***	0.264***
	(0.046)	(0.018)	(0.012)	(0.017)	(0.018)	(0.018)
Dumuy: New Construction	−0.901***	−0.881***	−0.883***	−0.861***	−0.881***	−0.883***
	(0.024)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Dumzny: Tracking	−0.106***	−0.103***	−0.106***	−0.107***	−0.101***	−0.107***
	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Dummy: Ground-mounted	0.372***	0.371***	0.359***	0.373***	0.361***	
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Dummy: Premium Module	0.088***	0.066***		0.064***	0.067***	0.053***
	(0.025)	(0.004)		(0.004)	(0.004)	(0.004)
Module Technology: Mono	0.069***	0.056***	0.047***	0.056***	0.056***	0.059***
	(0.015)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Module Technology; Other	0.127***	0.108***	0.088***	0.106***	0.108***	0.107***
	(0.023)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
MLPE: DC Optimizer	0.055***	0.059***	0.053***	0.062***	0.059***	0.060***
	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
ML.PE: None	−0.026***	−0.021***	−0.018***	−0.022***	−0.021***	−0.025***
	(0.006)	(0,003)	(0.003)	(0,003)	(0,003)	(0,003)
Electricity Price	0.850					
	(0.714)					
Households	−0.039***	−0.042***	−0.043***	−0,012***	−0.042***	−0.012***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
AGH/Aousehold	0.060***	0.061***	0.629***	0.060***	0.633***	
	(0.005)	(0.002)	(0.002)	(0.033)	(0.002)	(0.033)
Wages/Household				−0.610***		−0.607***
				(0.014)		(0.014)

(Continued)

Table 1—continued

Taxable Income/Household				0.067		0.051
				(0.039)		(0.039)
Income Tax/Houshold				−0.074***		−0.067***
				(0.018)		(0.019)
Constant	6.730***	8.630***	8.090***	8.650***	8.630***	8.310***
	(1.860)	(0.056)	(0.043)	(0.096)	(0.056)	(0.098)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	172,106	172,106	172,106	172,096	172,096	165,478
R^2	0.327	0,332	0.331	0.342	0.332	0.343
Adjusted R^2	0.327	0.332	0.331	0.342	0.332	0.343
Resadual Std. Error	0.410(df−172082)	0.409(df−172083)	0.409(df−172084)	0.406(df−172070)	0.409(df−172073)	0.403(df−165455)

Note: * p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001

Note: Second stage OLS regression results of explanatory variables on system size, using five different sets of instruments and four different sets of regressors. Model (3) is selected as the final model as it shows the highest R^2 and lowest standard errors among those models with no multicollinearity in the explanatory or control variables.

Source: Own analysis, estimation sample

percentage change from one group to the other (Benoit, 2011). A change in the respective variable changes the system size by $e^{\widehat{\beta}}$, or approximately by $\widehat{\beta}$ * 100%. Relative to polysilicon modules, using monocrystalline technology causes system size to increase by 4.7%. Likewise, the presence of both microinverters and DC optimizers is only weakly but positively related to system size installed. Intuitively, this might be related to the fact that installations with high-quality monocrystalline modules and MLPE technology became more and more prevalent, replacing less efficient systems (Barbose & Darghouth, 2019). Simultaneously, system size continuously increased (s. Appendix A2, Figure 7 and Figure 8). If the year fixed effects included in the model do not entirely capture this movement, the positive, though small coefficient might have resulted from this correlation.

Aside from efficiency-related aspects, the fact that a system is installed on a *new construction*, not as a retrofit on an existing house, strongly determines the system size installed, reducing it by 88.1%. The negative relationship could be due to the fact that new constructions are probably more space-constrained, offering less possibilities to install larger solar PV systems. Additionally, it should be taken into consideration that over 90% of the estimation sample are retrofits and that in the full sample, over 80% of the new construction installations were installed by a single company (Barbose & Darghouth, 2019). The estimate might, therefore, not offer an entirely valid representation.

Regarding the effect of *tracking technology*, the model states a 10.3% decrease in system size installed. As tracking is mainly used to improve the effective efficiency of a solar PV installation by maximising the amount of time the panels face the sun, this is probably particularly relevant for smaller systems with no space available to install further modules. On the contrary, installations that are *ground-mounted* show a 37.1% larger system size than rooftop systems. While the latter are constrained by a natural space limit of the house's roof, ground-mounted installations are likely to have a larger area available. Rooftop installations are also much more common among residential customers, accounting for over 95% of the estimation data.

For a percentage increase in the *electricity price* the model implies a 0.85% increase in system size (model (1)). The effect's direction is reasonable as higher electricity prices might induce customers to invest in more solar PV capacity in order to save money on the electricity bill or even earn some through net metering initiatives. However, this effect is not significant at a 5% level and the model exhibits high collinearity between numerous variables and electricity price, presumably because the information is aggregated and averaged on a state and year level. True electricity price movements may show substantial variation throughout the year or within a state.

Control variables

The effects' directions and sizes are consistent throughout the estimated models, using different sets of instruments and/ or different explanatory variables as regressors (s. Table 1). Some model specifications exhibit substantial multicollinearity, like for instance module efficiency and a dummy indicating a premium module in model (2) or several highly correlated control variables in models (4) and (6)[21]. As final model, I select the best specification in terms of a high coefficient of determination, a low residual standard error, low VIF values, and satisfactory diagnostic test results. Additionally, less complex models are preferable over more complex ones even if the bias of the estimates is lower for large models, because the variance increases in complexity (James et al., 2013). For these reasons, I refrain from including data on wages, population, taxable income, and income tax payment and limit the control variables to adjusted gross income per household and the number of households in a zip code area. Those are important to account for as they control for the average wealth and housing density in a given area. These aspects might influence the system size installed directly as well as through other factors included in the model. The preferred model estimates a small but negative elasticity of -0.04% upon a 1% increase in the *number of households* and a positive effect of 0.06% on system size resulting from a 1% increase in the *adjusted gross income per household*. Both estimate directions are plausible as higher income and more money available enable customers to invest in larger systems. Further, an increase in the number of households is often related to more urbanised regions where space is generally more limited, and installer costs might be higher due to larger market concentration. If more competition reduces installer experience and forces them to operate at smaller and less efficient scales (O'Shaughnessy, 2018), the coefficients might capture the effect of a resulting upward movement in prices compared to less densely populated regions, eventually decreasing the system size bought. Contrary to this, Gillingham et al. (2016) find that installer density is associated with substantially lower prices, likely due to reduced price mark-ups, which is more in line with common economic market theory (Mankiw, 2020; O'Shaughnessy, 2018).

Fixed effects

Besides explanatory and control variables the model includes fixed effects for the year and the U.S. state of installation. They are hard to interpret since fixed effects generally capture all influence over time and across states, respectively, that is not otherwise controlled for in the model. *Time fixed effects* are negative for all years throughout 2017 when compared to 2010 levels and only positive year-over-year for 2012 to 2013 (s. Appendix A6, Figure 24) despite the fact that median system size increased continuously (s. Appendix A6, Figure 7). One plausible reason for the negative

[21] I test all models on multicollinearity by computing the correlation matrices (s. Appendix A3, Figure 16 and Figure 17) and the VIF that indicates how much the variance of a coefficient is inflated (James et al., 2013) (s. Appendix A6, Table 13).

impact on installed capacity is that the fixed effects capture unobserved cost and price trends that are not included in the model. This might be the phase-out of incentive programs and rebates over the years, making systems less profitable (ceteris paribus). Cash rebates, for instance, are capacity-dependent, ending if a certain amount of solar PV has been installed in a given area (Consumer Energy Alliance, 2018; Shrimali & Jenner, 2013). On a federal level, the reduction of ITCs has been postponed to 2020, however. Furthermore, global negative demand shocks pushing up prices, policy changes and import tariffs, increasing wage rates, higher customer acquisition costs for more mature markets, and the price development of substitute sources of electricity like natural gas might be captured by the year coefficients (Burr, 2014; Newell & Raimi, 2014). As some of the factors named above change frequently, using more granular time effects on quarterly or weekly intervals might further improve the estimation. This will be part of the robustness checks in section 3.5.4.2 below.

State fixed effects capture differences across states, estimates showing a positive coefficient for Arizona and other states and a small but negative coefficient for Texas compared to California (s. Appendix A6, Figure 25). However, as there are only four distinct state groups in the model and California accounts for a major part of installations, these state fixed effects must be interpreted with caution. In terms of price-related aspects the effect sizes might be determined by the different implementation of various incentive programs like capacity- or performance-based compensation, local rebates, tax exemptions, and retail net metering compensation present in some U.S. states (Barbose & Darghouth, 2019; Burr, 2014). In California for instance, the California Solar Initiative offers capacity-based rebates, however with a declining absolute rebate amount per watt as system size increases (Burr, 2014). This may cause Californian investors to see a low marginal benefit in installing higher capacities, which is in line with the observation that California exhibits a systematically lower median system size than most other U.S. states (s. Appendix A2, Figure 8). Furthermore, BoS costs differ regionally as they depend more heavily on local wage rates, taxes and competition (Barbose & Darghouth, 2019). As they are not included as a model variable, these differences might be captured by the state coefficients. Apart from price and cost factors, local solar irradiance most likely plays another significant role as it varies across regions. Arizona is the sunniest state in the U.S. (NREL, 2020), increasing the profitability per kilowatts capacity installed compared to other states and, thus, incentivising customers to purchase larger systems. For California and Texas, the solar irradiance is similar on average though it differs within states and cities, suggesting that fixed effects on zip code or census block level might have further improved the estimation.

3.5.4. Robustness comparison to other model specifications

Varying the instruments

As already outlined above, I assess the predictive power of several instrument combinations to obtain the best possible approximation of the variation in solar PV installed prices. The results are displayed in Appendix A6. Particular importance is placed on the model quality measured by R^2 of the first stage as it is predictive for the accurate representation of price in the second stage estimation. Individual instruments as well as different instrument combinations are tested. For individual instruments, variation in the tax rate can explain 52.6% of the variation in price per watt, which is up to twice as much as all other instrument candidates (s. Appendix A6, Table 8). Adding polysilicon price movements as second instrument hardly improves the R^2 value (52.7%). Nevertheless, I use the combined set as instruments because polysilicon price movements capture the input factor cost side of the price development which might be more relevant and predictive in some model applications and time frames, also beyond the scope of this study. The first stage coefficient on polysilicon price is positive, price per watt increasing by 0.059% as polysilicon prices increase by 1%. Counterintuitively, with a percent-increase in the sales tax rate, the installed price per watt decreases. This can be explained by the fact that tax exemptions granted by the government are already regarded here. If more generous financial incentives like tax exemptions are granted in regions where prices are higher – which is in line with findings by Gillingham et al. (2016) – then the resulting effect of an increase in taxes can turn negative. For the data at hand this seems to be the case for some states (s. Appendix A6, Table 12). Further, it is not unreasonable to assume that there are numerous omitted variables taking effect here. Possible examples could be more subsidies granted to installers or lower BoS costs where tax rates are high. This needs to be investigated in more detail to identify the causation behind the estimated correlation.

Estimating the second stage as a simple linear model without further predictors the coefficient of determination[22] is low for all estimations, highlighting the need for further predictor variables (s. Appendix A6, Table 9). Therefore, numerous other model specifications are tested against the final model also in terms of instrumental variables used. Including the explanatory and control variables of the final model and accounting for year and state fixed effects strongly improves the predictive accuracy of the first stage, resulting in an R^2 of 0.992 for the final model (s. Appendix A6, Table 10). Nearly all the variation in the installed price per watt is now captured by the model. However, the polysilicon price coefficient in the first stage reduced form estimating the installed price is now very small and negative (-0.004), though still significant. This is counterintuitive as final prices are expected to move in the same direction as input factor

[22] Although R^2 is not a valid metric for IV regression estimates since "the actual values, not the instruments for the endogenous right-hand-side variables, are used to determine the model sum of squares" (Sribney, Wiggins, & Drukker, 2020), the value is also considered for evaluation as the correct manual computation of the second stage resulted in very similar R^2 values for all relevant model specifications.

costs. One possible explanation might be that time fixed effects now included in the model capture the effect of related changes in input costs that are correlated with the polysilicon price and had before been represented by its coefficient. This hypothesis is supported by a negligible Pearson correlation of price per watt and polysilicon price, only once year is held constant (s. Appendix A2, Figure 16 and Figure 17). Tax rate still has a negative and substantial impact on installed prices (-1.260). Adding further instruments improves the model only at the cost of high multicollinearity, if at all. Furthermore, diagnostic test results show endogeneity of instruments for some specifications[23]. On the other hand, using only the polysilicon price greatly reduces the R^2 to 0.188, in line with the findings from the simple linear first stage regression above. This shows that changes in polysilicon price do not suffice to capture the variation in the input factor costs. It might improve the predictive power to use variation in costs for intermediate products, as long as these variations can still be considered exogenous to the residential investment decision. Such production factors further down the value chain could be wafers or cells, and labour costs for technology development and system installation on a more granular distinction regarding region and time.

Fortunately, the predictive accuracy and good overall fit of the IV model – not the coefficients of the first stage – are of primary interest here. Therefore, a poor model specification in the first stage does not necessarily imply inconsistent coefficient estimates in the second stage as long as the fitted price values are accurate (Angrist & Krueger, 2001).

Varying the functional form

Next to instrument sets, I test four alternative model specifications against the final model described above as benchmark (Table 2, model (1)): A second-degree polynomial in the first stage, interaction of state and year fixed effects, quarterly time fixed effects, and no fixed effects.

The first alternative aims to obtain better prediction in the first stage. I estimate the price per watt with a *second-degree polynomial* for tax rate. Both coefficients are still highly significant, indicating that there might indeed be a non-linear relationship between prices and tax rate, also visible when investigating the regression plot of installed price on tax rate (s. Appendix A3, Figure 15). However, as the first stage coefficient of determination was already very high it could only be improved by 0.001 to 99.3% (s. Appendix A6, Table 11). Moreover, as mentioned above, the goodness of fit in the second stage does not depend on getting the first stage functional form exactly right (Angrist & Krueger, 2001; Kelejian, 1971), which is why the more complex first stage specification is not considered further.

Several legitimate reasons suggest that including an *interaction between state and year* might substantially improve the estimation by accounting for within-state differences over time, in effect allowing each state regression its own intercept and slope value (Borenstein et al., 2010). Contrary to expectations, the resulting elasticity estimates show hardly any difference, however. Presumably, the result would have turned out different for a broader range of data. With almost exclusively Californian installations from the years 2015 to 2017, the information available to detect state-specific differences over time is very limited. Notwithstanding, the first stage R^2 is slightly higher at 0.993 (s. Appendix A6, Table 11).

Quarterly fixed effects are tested against the baseline with yearly fixed effects to account for changes over shorter time intervals. The resulting estimates are displayed in Appendix A6, Figure 26. The second stage R^2 improves only marginally by 0.001, potentially because control data is only added on a yearly basis, offering no variation over quarters. I would generally expect more specific fixed effects to improve the estimation if there is reason to assume that there are differences between observational units. E.g., county, zip code, or even census block fixed effects could better capture constant regional differences in irradiance, community mindsets, and neighbourhood spillover.

Lastly, I use a model estimated *without fixed effects* as a simple benchmark to test whether they make a difference to the model fit. Surprisingly, the second stage R^2 is not much lower than for the other specifications, also when estimating it manually for the second stage. However, the first stage coefficient of determination diminishes (s. Appendix A6, Table 11), implying a worse fit of price per watt used in the subsequent step.

All in all, it is reassuring that the estimates do not differ much across most specifications which is why I choose the baseline, the simplest of the equally well performing models.

3.5.5. Impact of regional and economic differences

In the U.S., income inequality is a serious issue that is strongly linked to state welfare and policies (Jansa, 2020). I therefore group the observations in strata of income and population density as well as by state and estimate their price elasticities separately.

Income level impacts the price elasticity (Andreyeva, Long, & Brownell, 2010), although the effect has not been studied extensively, with most research rather focussing directly on income elasticities (Zhu, Li, Zhou, Zhang, & Yang, 2018). One expects consumers to be more price sensitive if they have little money to spend and vice versa (Mankiw, 2020). Running the preferred model regression for four different income groups[24] I find declining price elasticities as income level increases, meaning that households in zip codes with high average income have a substantially lower price elasticity (-0.195), i.e., are less price sensitive, than households with less

[23] Details on diagnostic tests are provided in section 3.5.6 'Tests on validity and model specification' for the final model.

[24] Observations are grouped by adjusted gross income per household. Equal intervals in thousands of USD are (1) low: (16.5,442], (2) low/medium: (442,867], (3) medium/high: (867,1.29e+03], (4) high: (1.29e+03,1.72e+03].

Table 2: IV regression results of robustness checks for the final model against alternative specifications

	Alternative specifications: Instrumental Variable Estimation Results				
	Dependent Variable: System Size (W)				
	IV: Baseline	IV: Polynomial in first stage	IV: Interaction state and year	IV: Quarterly fixed effects	IV: No fixed fixed effects
	(1)	(2)	(3)	(4)	(5)
Price per Watt	−0.443***	−0.446***	−0.443***	−0.443***	−0.385***
	(0.004)	(0,004)	(0,004)	(0.004)	(0.005)
Module Efficiency	0.408***	0.408***	0.409***	0.411***	0.243***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Dummy: New Construction	−0.883***	−0.884***	−0.884***	−0.870***	−0.858***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Dummy: Tracking	−0.106***	−0.106***	−0.104***	−0.105***	−0.111***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.013)
Dummy: Grotud-mounted	0.371***	0.371***	0.371***	0.371***	0.359***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Module Technology: Mono	0.047***	0.048***	0.047***	0.048***	0.052***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Module Technology: Other	0.088***	0.089***	0.088***	0.089***	0.101***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
MLPE: DC Optimizer	0.053***	0.053***	0.053***	0.054***	0.051***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
MLPE: None	−0.018***	−0.018***	−0.018***	−0.017***	0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Households	−0.043***	−0.043***	−0.043***	−0.043***	−0.046***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
AGL/Household	0.061***	0.061***	0.061***	0.060***	0.062***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	8.090***	8.090***	8.080***	8.100***	8.230***
	(0.043)	(0.043)	(0.043)	(0.046)	(0.043)
Time FE	Year	Year	Year	Quarter	None
State 11:	Yes	Yes	Yes	Yes	No
Interaction	No	No	Yes	No	No
Polynotaials	No	Yes	No	No	No
Cbservations	172,106	172,106	172,106	172,106	172,106
R^2	0.331	0.331	0.331	0.332	0.323
Adjusted R^2	0.331	0.331	0.331	0.332	0.323
Residual Std. Error	0.409	0.409	0.409	0.409	0.411
	(df = 172084)	(df = 172084)	(df = 172070)	(df = 172060)	(df = 172094)

Note: Second stage OLS regression results of explanatory variables on system size. The final model as baseline estimation compared to alternative specifications including a second-degree polynomial in the first stage, interaction of the fixed effects, quarterly time fixed effects, and no fixed effects. All use the same instruments polysilicon price and tax rate for price per watt.
Source: Own analysis, estimation sample

money available (-0.521) (s. Appendix A6, Table 14). These insights, though maybe unsurprising, are highly relevant in the context of targeted marketing and sales for price discrimination strategies as well as policy measures for tailored state subsidies and rebate campaigns.

The relevance of *population density* on price sensitivity is worth investigating as respective insights could help to focus firms' and governments' attention on regions where it is most profitable, both in an economic and societal welfare sense. The regression results displayed in Appendix A6, Table 15 for three groups of population density[25] show small but consistent differences throughout. Low population density regions

[25] Observations are grouped by the number of households. Equal intervals

are associated with a slightly higher price sensitivity (-0.473) than areas with high population density (-0.338). I control for income which might be correlated with the number of households in a zip code area, if the suburbs of metropolitan areas show systematically higher average income levels (Pendall & Carruthers, 2003; Schuetz et al., 2018).

I also run the regression separately for the states California, Arizona, and Texas, which are still included in the sample data when estimating the full model. Notably, the estimates for California hardly change compared to the full model as those made up most of the final observations. Coefficient estimates for Arizona and Texas deviate substantially for some variables, based on relatively few observations and partly missing variation in instrument data (s. Appendix A6, Table 16). This strongly suggests that statements can be made about California, but that the generalisability of the results to other Southern states let alone the entire U.S. should be considered with great caution.

3.5.6. Tests on validity and model specification

In order to ensure the sound application of the IV estimation approach, the validity of the instruments needs to be tested (Stock & Watson, 2020). This comprises two main aspects already outlined in section 3.2.2: Relevance and endogeneity. The diagnostic tests results for the final model are shown in Appendix A7, Table 17.

Relevance of the instruments can be assessed in a straightforward way by calculating the correlation coefficient between the instrument and the endogenous regressor it replaces (s. Appendix A3, Figure 16). The higher the correlation the more information is kept and the better its suitability as an instrument. On the contrary, weak instruments tend to mirror the OLS estimate (Angrist & Krueger, 2001; Angrist & Pischke, 2008). I use the first stage F-statistic test on weak instruments for the preferred model, effectively testing if coefficients on the instruments are all zero in the first stage. The resulting p-value of the test statistic is <0.001%. Hence, I can reject the hypothesis that the instruments are irrelevant.

Exogeneity of instruments is somewhat more complex to determine. It is only possible for an overidentified model that has more instruments than endogenous variables to be replaced[26]. Fortunately, this is the case here. The Sargan test (or J-Statistic) can be used to determine whether the residuals $\hat{u}^{TSLS}$ of the IV model can be explained by the instruments. If they cannot, one can assume that the instruments are indeed exogenous and valid for estimation. For the preferred model, the Sargan test statistic is insignificant. I cannot reject the hypothesis that the instruments show coefficients different from zero to explain the model residuals (Stock & Watson, 2020). Therefore, I can assume exogeneity of the instruments chosen here.

Additionally, I use the Wu-Hausman test to evaluate whether endogeneity is actually present in the original model (6) as otherwise OLS would be preferable over TSLS (Stock & Watson, 2020). The test statistic is again highly significant, and I can reject the hypothesis that OLS is consistent, supporting the TSLS approach used for estimation.

Apart from IV-related tests, I assess some general regression assumptions, outlined in Appendix B1. Corresponding plots are shown in Appendix A7, Figure 27 and Figure 28. The *linearity assumption* of the model seems valid for the first but not for the second stage. This implies that potentially a different type of model could improve the fit. This is not unreasonable as the final model selected here can only explain a rather small share of the variance in the installed system size. Plotting the residuals against the predicted system size values suggests that the *i.i.d. assumption* is violated and *heteroskedasticity* is present in both the first and the second stage. This is at least partly accounted for by using heteroskedasticity-robust standard errors in order to obtain valid test statistics and p-values. The *normality assumption* does not hold in the first stage and indicates heavy left and right tails in the second stage regression. However, due to the high sample size this is a minor issue here as the estimates converge asymptotically towards the true parameter as the number of observations increases (Ghasemi & Zahediasl, 2012; Lumley, Diehr, Emerson, & Chen, 2002; Stock & Watson, 2020).

4. Insight Relevance

4.1. Economic and Policy Implications

The U.S. are still one of the world's major emitters of greenhouse gases (EPA, 2020). In order to sustainably reduce emissions conscious consumption, energy efficiency measures, public policy encouragement, and, most importantly, the increased deployment and implementation of renewable energy sources are necessary (Jaforullah & King, 2015). How can the insights obtained in the analysis be leveraged to move further towards a more sustainable energy landscape in the U.S.? First of all, the knowledge of price elasticities and a more differentiated view on the sensitivity for various subgroups of the population can help to build more targeted and, thus, more effective and cost-efficient incentive programs. If in line with the general political directions, subsidies could be set higher for those who react stronger to price changes. Here, final cost reductions will promote an increase in solar PV capacity most. On the other hand, where price sensitivity is low, benefits like tax reductions could be decreased, generating higher governmental tax income which could ultimately be invested into initiatives where it has a potentially bigger impact. Besides the environmental benefit of more energy generated from renewable resources, the U.S. economy could profit substantially, also from the creation of new job opportunities for higher- and lower-skilled workers (Wei, Patadia, & Kammen, 2010).

Likewise, the results can be useful as quantitative evidence and guidance in the ongoing discussions on reducing

in thousands are (1) low: (0.09,18.1], (2) medium: (18.1,36.2], (3) high: (36.2,54.3].

[26] For an exactly identified model it is not possible to test exogeneity formally, but one must rely on logical reasoning (Stock & Watson, 2020).

or completely phasing out financial incentives (Gillingham & Tsvetanov, 2019a). From relative changes in installed capacity, policymakers can derive absolute capacity added. This allows forecasting changes in the overall capacity upon policy modifications using simple policy simulations and deploying the prior estimated pass-through rate and price elasticity (Gillingham & Tsvetanov, 2019a). Moreover, Gillingham and Tsvetanov (2019a) show that the potential reduction of GHG emissions from certain programs can be quantified. Naturally, the assumed amount of avoided emissions depends on the expectations on the type of generation that is displaced by renewables, both today and in future (Gillingham & Tsvetanov, 2019a). Gillingham and Tsvetanov (2019a) estimate the cost of abatement through state incentives for solar PV to lie notably above the U.S. government social costs of carbon estimate. Directly comparing the costs of the program to the social benefit of pollution abated allows to evaluate the cost-effectiveness and social desirability in an economic welfare sense. This method could be used to rank policy programs according to a quantifiable impact on social welfare and invest funds where they create the greatest benefit, also going beyond solar PV. This would take into account that a new technology should always be assessed not only in the light of cost-effectiveness but also environmental friendliness and social acceptance (Khan & Arsalan, 2016).

4.2. Business Implications

While governments can clearly leverage the insights on demand elasticities for solar PV in numerous ways, firms can also capitalise on them. Seeing that residential consumers are not very price sensitive, pricing strategies can be optimised to maximise profits, assuming that they operate in an imperfect market and do not set prices equal to marginal costs (Gillingham & Tsvetanov, 2019a; Mankiw, 2020). Besides that, more data-driven and cost-optimal decisions on research and development efforts to reduce costs even further can be taken. As the study also indicates different price sensitivity for population groups and regions, firms could consider targeted price discrimination. Likewise, the possibility to forecast the shift in consumption upon price changes more accurately also in the short-term enables firms to improve their production planning and draw near actual demand for a better supply-demand-fit.

5. Concluding Remarks and Outlook

5.1. Critical Review

Several aspects need to be mentioned when assessing the quality of the estimation results, regarding both data and methodology.

Data quality. The estimation results can only ever be as good as the quality of the data they are based on. The data in the TTS sample are self-reported by installers, depending on reporting conventions which potentially vary significantly. The scope of installed prices can sometimes even include warranties, monitoring and maintenance, re-roofing costs, and loan-related fees (Barbose & Darghouth, 2019). Additionally, installed prices do not necessarily reflect actual costs as they include profit margins and other installer-related characteristics which cannot be captured by cost-related instruments. What is more, according to Barbose and Darghouth (2019) the data set likely contains all kinds of systems, not only turnkey solutions, which is not fully visible from the system information and, thus, not considered in the estimation.

Data coverage. The data were only collected for some states mainly through incentive administrators (Barbose & Darghouth, 2019) which might cover most but probably not all installed systems in the U.S. This objection is further supported by a weak or missing sample coverage for smaller, often lower-cost, state markets. Self-selection bias could be present here (Heckman, 1990). Discarding observations with missing predictor values in the final data sample and assuming them to be missing at random can likewise lead to biased estimates, especially if installations with missing values differ systematically from the completely observed cases (Gelman & Hill, 2006). Furthermore, a broader data coverage is necessary for potential instrument data in order to optimally capture the variation in prices.

Data granularity. The sample of installations collected by the LNBL contains information on the installation date, making it possible to identify the timeframe of the decision to invest quite accurately, assuming that buyers base their decision on the most up-to-date information available at that point in time. The same holds true for very detailed location data, provided at zip code level. Unfortunately, most of the complementary data joined to the TTS sample were available only for a much broader time frame and region. This prevented the model from identifying variation on a more granular level which might have otherwise added significant information to the estimation and greatly improved its accuracy.

Besides those data-related aspects, I shed light on drawbacks of the estimation methodology and model specification.

Omitted variables. The most evident and pressing issue is the fact that by far not all relevant variables could be considered in the final model, either due to unavailability or unobservability of the information. For one thing, I did not consider the actual performance potential of individual systems, which is to a major part determined by the weather conditions and the amount of sunshine received in a certain installation location. Adding solar irradiance data on zip code or census block level could greatly improve the estimation. This is relevant for the initial decision to invest but might also influence the size of the system installed. Equally important, the initial installed price is only one determining factor to assess the economics of a PV installation. To gain comprehensive insights on the profitability and benefits, aspects such as ongoing operating and maintenance costs, effective performance, later retrofitting costs, saved electricity costs and payback period could also be considered as they are possibly already taken into account in the decision on how much capacity to invest in. Furthermore, both federal, state, and utility

support policies and regulations have hardly been taken into consideration but are most likely to have a tremendous effect on the buying decision. Capacity- and performance-based incentives will particularly impact the size of the system installed. Those data were not or only incompletely available and should be added for future estimations, if possible. Further, sociodemographic aspects, like investor age and education as well as idealistic values and mindset might impact the demand. The latter are unobservable but could be inferred from political party membership or voting behaviour (Iizuka, 2016; Matthew E. Kahn & Matsusaka, 1997). Additionally, Bollinger and Gillingham (2012), Graziano and Gillingham (2015), and Palm (2017) find significantly positive peer effects on the adoption of solar panels, suggesting that previous installations in the vicinity matter through neighbours attitudes and social influence. The omission of these variables, if they are correlated with any variable included in the model, will cause endogeneity and biased estimation results (Stock & Watson, 2020).

Model specification. The test on linearity in the second stage of estimating the log-log model suggests that the actual relationship might not be represented entirely accurately. It might be worth the effort to investigate which form better describes the relationship to improve the representation of the demand curve and obtain correct elasticity estimates. In case the price enters the model in a non-linear way or the model is even non-linear in its parameters (Imbens & Wooldridge, 2007; Wooldridge, 2015), control functions rather than standard IV methods should be applied, as done by Gillingham and Tsvetanov (2019a). Also, examining more profoundly which variables are highly relevant, differentiating between regions and population groups, would further back up the estimation. This could also help to identify even better instruments to further improve the first stage estimation.

5.2. Limitations

As with most research, the estimation results in this study are highly context-sensitive. Consequently, their application to other contexts is neither straight-forward nor generally possible. Although the goal of the study is to estimate the PV demand and price elasticity for the entire U.S., the final data sample contains almost exclusively Californian installations. Therefore, generalising these estimation results even to other states, let alone countries or cultures, is not unreservedly recommended, but must be done under consideration of numerous relevant aspects, if at all. Additionally, the estimation sample is limited to residential customers. There is good reason to assume that estimates would be substantially different in non-residential and utility-scale applications (Barbose & Darghouth, 2019). Furthermore, as pointed out several times before, the price sensitivity is estimated regarding the system size demanded rather than the discrete adoption decision. This needs to be taken into consideration not only when interpreting the coefficient estimates but also when generalising to a broader scope as the price elasticities regarding the system size are only accurate given that a consumer decided to invest in solar PV.

Lastly, one needs to be aware that correlation does not imply causation (Altman & Krzywinski, 2015; Holland, 1986). More than often, there are multiple ways to interpret coefficients. As a regression is unable to picture a causal direction, one needs to bear in mind that equations build on assumptions regarding the underlying causality (Altman & Krzywinski, 2015). If I do not entirely account for endogeneity in the model, variation in some variables might still bias coefficient estimates. This raises the question whether it is justifiable to report ceteris paribus effects.

5.3. Outlook and Conclusion

The study provides methodological insights as well as practical recommendations. Thoroughly and holistically addressing the concerns outlined above would be another big step forward towards applying the model not only to solar photovoltaics but also to many other recent and emerging technologies (Gillingham & Tsvetanov, 2019a, 2019b). This can yield valuable insights for future governmental policies and firm decisions alike, potentially also providing a way to assess the effectiveness of different types of programs. Further, extending the estimation to the technology adoption as explained variable in a two-stage model would enable more far-reaching and differentiated statements (e.g., Bollinger & Gillingham, 2012; Cui, 2018; Dong & Sigrin, 2019; Gillingham & Tsvetanov, 2019a; Palm, 2017; Rogers & Sexton, 2014).

Understanding the factors determining the propensity to invest in solar PV is particularly relevant when it comes to demand forecasting. The within-sample insights obtained in this study could be leveraged in order to make out-of-sample predictions. Numerous machine learning techniques are well-applicable to the estimation problem at hand. Especially non-parametric, tree-based learners like Random Forest and Gradient Boosting have proven highly beneficial for both classification and regression in many economic and business applications (James et al., 2013; Murdoch, Singh, Kumbier, Abbasi-Asl, & Yu, 2019). At the expense of interpretability (Orrù, Monaro, Conversano, Gemignani, & Sartori, 2020), they can use huge amounts of data and features while making no assumption on the functional form, which is convenient when the model is complex. Furthermore, non-parametric learners natively handle outliers and multicollinearity and are able to capture regional differences and non-linear relationships, which is highly relevant in case location matters (James et al., 2013). Using these methods on the estimation problem at hand could greatly supplement the interpretable insights obtained in this study and increase its relevance for researchers, businesses, as well as state and federal governments.

For now, the study successfully provides insights on the price elasticity for solar PV – if not for the entire U.S., at least for California. It shows that customers are generally rather insensitive to price changes. It also brings to light the relevance of other factors impacting the demand, module efficiency being almost as relevant as price per watt when it comes to the system size installed. The comparison of price

elasticities within various population subsets highlights the potential of more targeted interventions to maximise the energy amount generated from renewable resources and promote the reduction of greenhouse gas emissions. Thereby, solar photovoltaics can indeed make a major contribution to the sustainable transformation of the energy and electricity generation landscape in the United States. Setting sound policies and incentive programs based on the findings described above, America could even attain to become what Barack Obama aspired to seven years ago: A leader in the global transition towards a sustainable economy.

References

Altman, N., & Krzywinski, M. (2015). Association, correlation and causation [Journal Article]. *Nature Methods*, *12*(10), 899-900. Retrieved from https://doi.org/10.1038/nmeth.3587 doi: 10.1038/nmeth.3587

Andreyeva, T., Long, M. W., & Brownell, K. D. (2010). The impact of food prices on consumption: A systematic review of research on the price elasticity of demand for food [Journal Article]. *American Journal of Public Health*, *100*(2), 216-222. Retrieved from https://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2008.151415 doi: 10.2105/ajph.2008.151415

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments [Journal Article]. *Journal of Economic perspectives*, *15*(4), 69-85.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion* [Book]. Princeton University Press.

Baltagi, B. (2008). *Econometric analysis of panel data* [Book]. John Wiley & Sons.

Barbose, G., & Darghouth, N. (2019). *Tracking the sun: Pricing and design trends for distributed photovoltaic systems in the united states, 2019 edition* (Report). Lawrence Berkeley National Laboratory, Berkeley, USA. Retrieved from https://emp.lbl.gov/tracking-the-sun

Baum, C., Schaffer, M., & Stillman, S. (2007). *Ivendog: Stata module to calculate durbin-wu-hausman endogeneity test after ivreg* [Computer Program]. Retrieved from https://EconPapers.repec.org/RePEc:boc:bocode:s429401

Benoit, K. (2011). Linear regression models with logarithmic transformations [Journal Article]. *London School of Economics, London*, *22*(1), 23-36.

Bernstein, M. A., & Griffin, J. (2006). *Regional differences in the price-elasticity of demand for energy* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.nrel.gov/docs/fy06osti/39512.pdf

Blakers, A. (2019). Development of the perc solar cell [Journal Article]. *IEEE Journal of Photovoltaics*, *9*(3), 629-635. doi: 10.1109/JPHOTOV.2019.2899460

Bollinger, B., & Gillingham, K. (2012). Peer effects in the diffusion of solar photovoltaic panels [Journal Article]. *Marketing Science*, *31*(6), 900-912. Retrieved from https://EconPapers.repec.org/RePEc:inm:ormksc:v:31:y:2012:i:6:p:900-912

Bollinger, B., & Gillingham, K. (2019). Learning-by-doing in solar photovoltaic installations [Journal Article]. *Available at SSRN 2342406*. doi: http://dx.doi.org/10.2139/ssrn.2342406

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis [Journal Article]. *Research synthesis methods*, *1*(2), 97-111.

Burr, C. (2014). Subsidies, tariffs and investments in the solar power market [Journal Article]. *University of Colorado-Boulder Working Paper*.

Center for Climate and Energy Solutions. (2019). *Paris climate agreement q&a* (Report). Retrieved from https://www.c2es.org/site/assets/uploads/2017/09/paris-climate-agreement-qa.pdf

Center for Sustainable Systems. (2019). *U.s. energy system factsheet* (Report). University of Michigan, Ann Arbor, USA. Retrieved from http://css.umich.edu/sites/default/files/US%20Energy%20System_CSS03-11_e2019.pdf

Consumer Energy Alliance. (2018). *Incentivizing solar energy: An in-depth analysis of u.s. solar incentives* (Report). Retrieved from https://consumerenergyalliance.org/cms/wp-content/uploads/2018/06/Solar-incentive-report-060418.pdf

Cui, M. (2018). Demand and supply for rooftop solar [Conference Proceedings]. In *University of california, los angeles, usa.*

Deline, C., Meydbray, J., Donovan, M., & Forrest, J. (2012). *Photovoltaic shading testbed for module-level power electronics* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.osti.gov/servlets/purl/1045715 doi: https://dx.doi.org/10.2172/1045715

Denholm, P., & Margolis, R. (2007). *Regional per capita solar electric footprint for the united states* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.nrel.gov/docs/fy08osti/42463.pdf

Dennis, B. (2019, 2020, April 9). [Electronic Article]. Retrieved from https://www.washingtonpost.com/climate-environment/2019/11/04/trump-makes-it-official-us-will-withdraw-paris-climate-accord/

Dong, C., & Sigrin, B. (2019). Using willingness to pay to forecast the adoption of solar photovoltaics: A "parameterization + calibration" approach [Journal Article]. *Energy Policy*, *129*, 100-110.

Dong, C., Wiser, R., & Rai, V. (2018). Incentive pass-through for residential solar systems in california [Journal Article]. *Energy Economics*, *72*, 154-165. Retrieved from http://www.sciencedirect.com/science/article/pii/S0140988318301397 doi: https://doi.org/10.1016/j.eneco.2018.04.014

EIA. (2019a). *Annual energy outlook 2019* (Report). U.S. Energy Information Administration, Washington DC, USA. Retrieved from https://www.eia.gov/outlooks/aeo/

EIA. (2019b). *Monthly energy review (april 2019)* (Report). U.S. Energy Information Administration, Washington DC, USA. Retrieved from https://www.eia.gov/totalenergy/data/monthly/

EIA. (2019c). (Vol. 2020,) (Web Page No. April 4). Retrieved from https://www.eia.gov/todayinenergy/detail.php?id=39092

EIA. (2019d). (Vol. 2020,) (Web Page No. July 25). Retrieved from https://www.eia.gov/todayinenergy/detail.php?id=40873

EIA. (2020a). *Annual energy outlook 2020* (Report). U.S. Energy Information Administration, Washington DC, USA. Retrieved from https://www.eia.gov/outlooks/aeo/

EIA. (2020b). *Monthly energy review (june 2020)* (Report). U.S. Energy Information Administration, Washington DC, USA. Retrieved from https://www.eia.gov/totalenergy/data/monthly/

Elshurafa, A. M., Albardi, S. R., Bigerna, S., & Bollino, C. A. (2018). Estimating the learning curve of solar pv balance–of–system for over 20 countries: Implications and policy recommendations [Journal Article]. *Journal of Cleaner Production*, *196*, 122-134. Retrieved from http://www.sciencedirect.com/science/article/pii/S0959652618316652 doi: https://doi.org/10.1016/j.jclepro.2018.06.016

EPA. (2020). *Inventory of u.s. greenhouse gas emissions and sinks: 1990-2018* (Report). U.S. Environmental Protection Agency, Washington DC, USA. Retrieved from https://www.epa.gov/sites/production/files/2020-04/documents/us-ghg-inventory-2020-main-text.pdf

Feldman, D. J., & Margolis, R. M. (2020). *Q4 2019/q1 2020 solar industry update* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.nrel.gov/docs/fy20osti/77010.pdf

Friedman, L. (2019, 2020, April 9). [Electronic Article]. Retrieved from https://www.nytimes.com/2019/11/04/climate/trump-paris-agreement-climate.html

Frondel, M., & Vance, C. (2013). Fuel taxes versus efficiency standards–an instrumental variable approach [Journal Article]. *Ruhr Economic Paper*(445).

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models* [Book]. Cambridge: Cambridge University Press. Retrieved from https://www.cambridge.org/core/books/data-analysis-using-regression-and-multilevelhierarchical-models/32A29531C7FD730C3A68951A17C9D983 doi: DOI:10.1017/CBO9780511790942

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians [Journal Article]. *International journal of endocrinology and metabolism*, *10*(2), 486.

Gillingham, K., & Bollinger, B. (2019). Social learning and solar photovoltaic adoption [Journal Article]. *Available at SSRN 3500930*. doi: https://dx.doi.org/10.2139/ssrn.3500930

Gillingham, K., Deng, H., Wiser, R., Darghouth, N., Nemet, G., Barbose, G., ... Dong, C. (2016). Deconstructing solar photovoltaic pricing [Journal Article]. *The Energy Journal*, *37*(3).

Gillingham, K., & Tsvetanov, T. (2019a). Hurdles and steps: Estimating demand for solar photovoltaics [Journal Article]. *Quantitative Eco-*

nomics, *10*(1), 275-310.

Gillingham, K., & Tsvetanov, T. (2019b). Supplement to "hurdles and steps: Estimating demand for solar photovoltaics" [Journal Article].

Graziano, M., & Gillingham, K. (2015). Spatial patterns of solar photovoltaic system adoption: the influence of neighbors and the built environment [Journal Article]. *Journal of Economic Geography*, *15*(4), 815-839.

Heckman, J. J. (1990). Selection bias and self-selection [Book Section]. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Econometrics* (p. 201-224). London: Palgrave Macmillan UK. Retrieved from https://doi.org/10.1007/978-1-349-20570-7_29 doi: 10.1007/978-1-349-20570-7_29

Heckman, J. J., & Navarro, S. (2007). Dynamic discrete choice and dynamic treatment effects [Journal Article]. *Journal of Econometrics*, *136*(2), 341-396. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304407605002186 doi: https://doi.org/10.1016/j.jeconom.2005.11.002

Holland, P. W. (1986). Statistics and causal inference [Journal Article]. *Journal of the American Statistical Association*, *81*(396), 945-960. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478354 doi: 10.1080/01621459.1986.10478354

Hughes, J., & Podolefsky, M. (2015). Getting green with solar subsidies: Evidence from the california solar initiative [Journal Article]. *Journal of the Association of Environmental and Resource Economists*, *2*(2), 235-275. Retrieved from https://EconPapers.repec.org/RePEc:ucp:jaerec:doi:10.1086/681131

IEA. (2019). *Renewables 2019* (Report). International Energy Agency, Paris, France. Retrieved from https://www.iea.org/reports/renewables-2019

Iizuka, M. (2016). Role of environmental awareness in achieving sustainable development [Journal Article].

Imbens, G., & Wooldridge, J. (2007). What's new in econometrics? lecture 6: Control functions and related methods [Journal Article]. *NBER*, *6*, 420-445. Retrieved from https://www.nber.org/WNE/Slides7-31-07/Slides7-31-07.pdf

IRENA. (2019). *Renewable power generation cost in 2018* (Report). International Renewable Energy Agency, Abu Dhabi, United Arab Emirates. Retrieved from https://www.irena.org/publications/2019/May/Renewable-power-generation-costs-in-2018

Jaforullah, M., & King, A. (2015). Does the use of renewable energy sources mitigate co2 emissions? a reassessment of the us evidence [Journal Article]. *Energy Economics*, *49*, 711-717. Retrieved from http://www.sciencedirect.com/science/article/pii/S0140988315001280 doi: https://doi.org/10.1016/j.eneco.2015.04.006

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112) [Book]. Springer.

Jansa, J. M. (2020). Chasing disparity: Economic development incentives and income inequality in the u.s. states [Journal Article]. *State Politics & Policy Quarterly*, *0*(0), 1532440019900259. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/1532440019900259 doi: 10.1177/1532440019900259

Kelejian, H. H. (1971). Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables [Journal Article]. *Journal of the American Statistical Association*, *66*(334), 373-374.

Keyes, A. T., Lambert, K. F., Burtraw, D., Buonocore, J. J., Levy, J. I., & Driscoll, C. T. (2019). The affordable clean energy rule and the impact of emissions rebound on carbon dioxide and criteria air pollutant emissions [Journal Article]. *Environmental Research Letters*, *14*(4), 044018.

Khan, J., & Arsalan, M. H. (2016). Solar power technologies for sustainable electricity generation – a review [Journal Article]. *Renewable and Sustainable Energy Reviews*, *55*, 414-425. Retrieved from http://www.sciencedirect.com/science/article/pii/S1364032115012149 doi: https://doi.org/10.1016/j.rser.2015.10.135

Krupnick, A., Cropper, M., Kopp, R., Newell, R., Pizer, W. A., Rennert, K., ... Keyes, A. (2018). Comments to us epa on the proposed affordable clean energy rule [Journal Article]. *Washington, DC: Resources for the Future*.

Lin, M., Lucas Jr, H. C., & Shmueli, G. (2013). Too big to fail: Large samples and the p-value problem [Journal Article]. *Information Systems Research*, *24*(4), 906-917.

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets [Journal Article]. *Annual review of public health*, *23*(1), 151-169.

MacKay, A., & Miller, N. (2019). Estimating models of supply and demand: Instruments and covariance restrictions [Journal Article]. *Available at SSRN 3025845*. doi: https://dx.doi.org/10.2139/ssrn.3025845

Mankiw, N. G. (2020). *Principles of economics* [Book]. Cengage Learning.

Matthew E. Kahn, B., & Matsusaka, J. G. (1997). Demand for environmental goods: Evidence from voting patterns on california initiatives [Journal Article]. *The Journal of Law and Economics*, *40*(1), 137-174. Retrieved from https://www.journals.uchicago.edu/doi/abs/10.1086/467369 doi: 10.1086/467369

Mewton, R. T., & Cacho, O. J. (2011). Green power voluntary purchases: Price elasticity and policy analysis [Journal Article]. *Energy Policy*, *39*(1), 377-385.

Miller, M., & Alberini, A. (2016). Sensitivity of price elasticity of demand to aggregation, unobserved heterogeneity, price trends, and price endogeneity: Evidence from us data [Journal Article]. *Energy Policy*, *97*, 235-249.

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Interpretable machine learning: definitions, methods, and applications [Journal Article]. *arXiv preprint arXiv:1901.04592*.

Newell, R. G., & Raimi, D. (2014). Implications of shale gas development for climate change [Journal Article]. *Environmental science & technology*, *48*(15), 8360-8368.

NREL. (2015, 2015-09-01). *Building america case study: Photovoltaic systems with module-level power electronics* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.osti.gov/servlets/purl/1225312

NREL. (2020, August 8, 2020). [Online Database]. Retrieved from https://www.nrel.gov/gis/solar.html

Nugent, D., & Sovacool, B. K. (2014). Assessing the lifecycle greenhouse gas emissions from solar pv and wind energy: A critical meta-survey [Journal Article]. *Energy Policy*, *65*, 229-244.

Orrù, G., Monaro, M., Conversano, C., Gemignani, A., & Sartori, G. (2020). Machine learning in psychometrics and psychological research [Journal Article]. *Frontiers in Psychology*, *10*(2970). Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2019.02970 doi: 10.3389/fpsyg.2019.02970

O'Shaughnessy, E. J. (2018). *The effects of market concentration on residential solar pv prices: Competition, installer scale, and soft costs* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.osti.gov/biblio/1452704 doi: 10.2172/1452704

Palm, A. (2017). Peer effects in residential solar photovoltaics adoption—a mixed methods study of swedish users [Journal Article]. *Energy Research & Social Science*, *26*, 1-10.

Parida, B., Iniyan, S., & Goic, R. (2011). A review of solar photovoltaic technologies [Journal Article]. *Renewable and Sustainable Energy Reviews*, *15*(3), 1625-1636. Retrieved from http://www.sciencedirect.com/science/article/pii/S1364032110004016 doi: https://doi.org/10.1016/j.rser.2010.11.032

Pendall, R., & Carruthers, J. I. (2003). Does density exacerbate income segregation? evidence from u.s. metropolitan areas, 1980 to 2000 [Journal Article]. *Housing Policy Debate*, *14*(4), 541-589. Retrieved from https://doi.org/10.1080/10511482.2003.9521487 doi: 10.1080/10511482.2003.9521487

Perea, A., Smith, C., Davis, M., Sun, X., White, B., Cox, M., ... Baca, J. (2020a). *U.s. solar market insight report 2019 year in review* (Report). Solar Energy Industries Association, Washington DC, USA. Retrieved from https://www.seia.org/research-resources/solar-market-insight-report-2019-year-review

Perea, A., Smith, C., Davis, M., Sun, X., White, B., Cox, M., ... Baca, J. (2020b). *U.s. solar market insight report q2 2020* (Report). Solar Energy Industries Association, Washington DC, USA. Retrieved from https://www.woodmac.com/research/products/power-and-renewables/us-solar-market-insight/

Platzer, M. D. (2012). Us solar photovoltaic manufacturing: Industry trends, global competition, federal support [Conference Proceedings]. In *Library of congress, congressional research service.* Library of Congress, Congressional Research Service.

Raman, S. (2013). Fossilizing renewable energies [Journal Article]. *Science as Culture*, *22*(2), 172-180.

Rogers, E., & Sexton, S. (2014). Effectiveness of subsidies for residential rooftop solar adoption: Lessons from california [Journal Article]. *North Carolina State University Working Paper*.

Schuetz, J., Larrimore, J., Merry, E. A., Robles, B. J., Tranfaglia, A., & Gonzalez, A. (2018). Are central cities poor and non-white? [Journal Article]. *Journal of Housing Economics*, *40*, 83-94.

SEIA. (2019). *Section 201 solar tariffs* (Report). Solar Energy Industries Association, Washington DC, USA. Retrieved from https://www.seia.org/sites/default/files/2019-12/SEIA-Section-201-Factsheet-Dec2019.pdf

SEIA. (2020). *Solar investment tax credit* (Report). Solar Energy Industries Association, Washington DC, USA. Retrieved from https://www.seia.org/sites/default/files/2020-01/SEIA-ITC-Factsheet-2020-Jan_1.pdf

Shrimali, G., & Jenner, S. (2013). The impact of state policy on deployment and cost of solar photovoltaic technology in the u.s.: A sector-specific empirical analysis [Journal Article]. *Renewable Energy*, *60*, 679-690. Retrieved from http://www.sciencedirect.com/science/article/pii/S0960148113003169 doi: https://doi.org/10.1016/j.renene.2013.06.023

Singh, G. K. (2013). Solar power generation by pv (photovoltaic) technology: A review [Journal Article]. *Energy*, *53*, 1-13. Retrieved from http://www.sciencedirect.com/science/article/pii/S0360544213001758 doi: https://doi.org/10.1016/j.energy.2013.02.057

Sribney, W., Wiggins, V., & Drukker, D. (2020). (Vol. 2020,) (Web Page No. August 2). Retrieved from https://www.stata.com/support/faqs/statistics/two-stage-least-squares/

Stock, J. H., & Watson, M. W. (2020). *Introduction to econometrics* [Book]. Pearson.

Tax Policy Center. (2020). *The state of state (and local) tax policy* (Report). Urban-Brookings Tax Policy Center, Washington DC, USA. Retrieved from https://www.taxpolicycenter.org/sites/default/files/briefing-book/how_do_state_and_local_sales_taxes_work_1.pdf

United Nations. (2016, 2020, April 16). [Online Database]. Retrieved from https://www4.unfccc.int/sites/NDCStaging/Pages/All.aspx

van der Hulst, M. K., Huijbregts, M. A., van Loon, N., Theelen, M., Kootstra, L., Bergesen, J. D., & Hauck, M. (2020). A systematic approach to assess the environmental impact of emerging technologies: A case study for the ghg footprint of cigs solar photovoltaic laminate [Journal Article]. *Journal of Industrial Ecology*.

Wei, M., Patadia, S., & Kammen, D. M. (2010). Putting renewables and energy efficiency to work: How many jobs can the clean energy industry generate in the us? [Journal Article]. *Energy Policy*, *38*(2), 919-931. Retrieved from http://www.sciencedirect.com/science/article/pii/S0301421509007915 doi: https://doi.org/10.1016/j.enpol.2009.10.044

Woodhouse, M., Smith, B., Ramdas, A., & Margolis, R. (2019). *Crystalline silicon photovoltaic module manufacturing costs and sustainable pricing: 1h 2018 benchmark and cost reduction road map* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.nrel.gov/docs/fy19osti/72134.pdf

Wooldridge, J. M. (2005). Instrumental variables estimation with panel data [Journal Article]. *Econometric Theory*, *21*(4), 865-869. Retrieved from www.jstor.org/stable/3533399

Wooldridge, J. M. (2015). Control function methods in applied econometrics [Journal Article]. *Journal of Human Resources*, *50*(2), 420-445.

Wright, P. G. (1928). *Tariff on animal and vegetable oils* [Book]. Macmillan Company, New York.

Yin, H., & Powers, N. (2010). Do state renewable portfolio standards promote in-state renewable generation? [Journal Article]. *Energy Policy*, *38*(2), 1140-1149.

Zainudin, H. N., & Mekhilef, S. (2010). Comparison study of maximum power point tracker techniques for pv systems [Conference Proceedings]. In *14th international middle east power systems conference, cairo university, egypt*.

Zhang, H.-B., Dai, H.-C., Lai, H.-X., & Wang, W.-T. (2017). Us withdrawal from the paris agreement: Reasons, impacts, and china's response [Journal Article]. *Advances in Climate Change Research*, *8*(4), 220-225.

Zhu, X., Li, L., Zhou, K., Zhang, X., & Yang, S. (2018). A meta-analysis on the price elasticity and income elasticity of residential electricity demand [Journal Article]. *Journal of Cleaner Production*, *201*, 169-177. Retrieved from http://www.sciencedirect.com/science/article/pii/S0959652618323588 doi: https://doi.org/10.1016/j.jclepro.2018.08.027

Junior Management Science 7(3) (2022) 668-689

Junior Management Science

journal homepage: www.jums.academy

Analysis of Green Bonds

Tobias Friedrich Fauß

Eberhard Karls Universität Tübingen & University of Nottingham

Abstract

Issuing its first green federal security in 2020, Germany pioneered a unique twin bond concept to address potential liquidity risks compared to their conventional counterparts. A switch mechanism between green and conventional bonds was introduced that allows debt-neutral sale-and-purchase (switch) transactions by the issuing authority. The main goal of this dissertation is to provide a theoretical model that is capable to explain the effects of this twin bond concept on the pricing of green bonds. For this purpose, a stochastic liquidity premium following a Vasicek (1977) process, a constant green premium and a switch option, which is executed when the green bond price falls below the price of its conventional twin bond, are assumed. The model results confirm that this twin bond concept is a viable solution to mitigate illiquidity-induced costs for the green bonds. The main learning from the model is a potential positive value of the switch option before its execution. This implies that issuers adopting this concept could benefit from lower costs of capital compared to ordinary green bonds without a switch mechanism. For investors holding the green instruments, this implies a reduced exposure to liquidity risks.

Keywords: Green bonds; German twin bonds; green premium; liquidity premium; switch transactions.

1. Introduction

Most nations have acknowledged the risks of climate change and pledged to pursue mitigating measures. As of today, 193 Parties adopted the 2015 Paris Agreement on climate change with the commitment to limit global temperature increase to 1.5°C above pre-industrial levels, and all United Nations Member States committed to the 17 Sustainable Development Goals (United Nations, n.d., 2022). This transition to a more sustainable global economy requires a substantial amount of new investments. For example, the European Commission anticipates additional annual investment needs of approximately 2.3% of GDP (i.e., 336 bn. EUR) for necessary energy system investments (exclusive transport) in light of its 2030 Climate Plan and 1.6% of GDP thereafter, aiming to become climate neutral by 2050 (European Commission, 2020). Similarly, the German Federal Government may need to increase its annual expenses from 200 bn. EUR to about 240 bn. EUR to become climate neutral until 2045, which is an additional percentage point of its 2019 GDP, cumulating to approximately 7% of GDP in total (Helmcke, Heuss, Hieronimus, & Engel, 2021). In addition to funding from the public sector, private investments can play a crucial role to provide the required financial resources (European Commission, 2020).

One instrument to raise funding for this transition is the emission of green bonds. Such fixed income debt securities differ from conventional bonds as their proceeds are entirely dedicated to the financing of environmental or climate related projects (Ehlers & Packer, 2017). Ehlers and Packer (2017) show that there is no single global definition for projects that fall into this category, but a range of different established standards and external verification procedures instead. For example, one widely accepted industry standard that is also adopted by the German Green Bond Framework are the Green Bond Principles issued by the International Capital Market Association (ICMA) (Finanzagentur GmbH, 2020). Other external validation concepts are second-party opinions by independent research institutes such as the Center for International Climate Research (CICERO), verification by auditors such as KPMG, certifications by organizations such as the Climate Bond Initiative (CBI) or green ratings, by rating agencies such as Standard & Poor's and Moody's (Dorfleitner, Utz, & Zhang, 2021). Finally, there also exist regional standards such as the EU Green Bond Standard or China's Green Bond Endorsed Project Catalogue (European Commission, 2021; People's Bank of China, 2021).

DOI: https://doi.org/10.5282/jums/v7i3pp668-689

The potential benefit of the above-outlined references is the reduction of asymmetric information and increase of transparency (Dorfleitner et al., 2021). The relevance of such additional information is supported by a CBI (2019) survey, which identified green credentials as one of the main drivers for green investment decisions.

The global market for green bonds has strongly increased over the past years. Based on figures published by the CBI (2021), their issuance volume already surpassed the former annual maximum of 2020 (i.e., 294.4 billion (bn.) US Dollar (USD)) in Q3 of 2021, with 354.2 bn. USD Year to Date (YTD). Further, they forecast the annual volume to exceed one trillion USD by 2023. To put these numbers into perspective, SIFMA (2021) reports for 2020 a global long-term bond issuance volume of 27.3 trillion (tn.) USD, indicating a green bond market share of about 1% that year.

While green bonds can help to finance sustainable investments, they may also incur additional expenses for their issuers compared to conventional bonds. These can be caused by internal costs due to screening, managing and reporting on their use of proceeds, as well as external costs for their certification and second-party opinion (Hachenberg & Schiereck, 2018). However, these additional costs might be compensated if investors require a lower return (i.e., yield to maturity) for holding the green assets. In terms of prices, a lower yield means that a green bond can be issued at a higher price in comparison to a conventional twin bond with the same nominal value, which reduces funding costs. In fact, evidence for lower yields of green bonds is found by the majority of empirical studies (MacAskill, Roca, Liu, Stewart, & Sahin, 2021). This implies that in spite of additional expenditures, issuers could even benefit from lower financing costs for their sustainable investments by issuing green bonds instead.

One characteristic that potentially influences the yield of green bonds is their degree of liquidity (e.g., Zerbib (2019)). This is because investors may require a higher return for holding an illiquid asset (Kempf & Uhrig-Homburg, 2000). The German Finance Agency (Finanzagentur) (FA), which is responsible for the issuance of German green sovereign bonds, argues that an excessive volume of green bonds may impede the liquidity of conventional bonds, while a deficient volume may impede their own liquidity (Finanzagentur GmbH, 2021a). In light of this trade-off, they pioneered a unique green bond concept to solve this issue in 2020, which in 2022 was adopted by Denmark as well (Dutch State Treasury Agency, 2019; Finanzagentur GmbH, 2020). In summary, this concept bases on the issuance of green bonds as twins to conventional bonds that coincide in almost all characteristics. This allows the introduction of a switch mechanism between both twins that has the function to secure a superior value of green bonds, which differ to conventional twins in its more restricted use of proceeds. Or in terms of yields, the yield of conventional bonds can serve as an upper limit for the yield of green bonds. For investors, this can imply additional certainty to sell a green bond for at least the price of its conventional counterpart. For issuers, this may imply more favourable financing conditions, as the green bonds can possess a lower yield.

The goal of this dissertation is the derivation of a theoretic model that is capable to explain the mechanisms of the German green bond concept. In detail, it aims to provide a decomposition of the yield differential between both twins into its individual components. Namely, a green premium, a liquidity premium and the added value of the switch mechanism. A successful disentanglement of the yield differential can provide issuers as well as investors with crucial information for evaluating this concept. From an issuers' perspective, this may answer the question if the framework is a viable approach to mitigate undesired illiquidity-induced costs and thus secure more favourable costs of capital. From an investors' perspective, this information can also be relevant to correctly account for their exposure to potential liquidity risks. The added value of this dissertation is therefore viewed as a theoretical contribution to improve the understanding of the implications of the German twin bond approach with a focus on its most defining feature, the switch mechanism between green and conventional twin bonds.

The remainder of this paper is structured as follows. First, an overview of relevant literature is provided. Then an evaluation of the German green bond concept and a comparison to their green sovereign peer bonds is conducted. In the following part, the theoretical model is derived and a calibration of the model parameters is performed. Finally, the model implications are evaluated, including a sensitivity analysis and a discussion of its limitations.

2. Literature Review

2.1. Green Premium

There exists an increasing body of literature with the goal to explain and quantify the potential yield premium for green bonds. Such a green premium can be defined as the difference in yield to a conventional bond, which shares otherwise the same characteristics (Zerbib, 2019). A negative premium implies that investors require a lower return when investing into a green asset.

Fama and French (2007) find that the taste for an asset, expressed by additional utility from holding it beyond its financial payoff, can help to explain its prices. In line with this result, Dorfleitner et al. (2021) argue in the context of green bonds that a lower yield for holding a green asset may be compensated by a non-financial utility component. This is also supported by findings from Riedl and Smeets (2017) who observe that social preferences and signalling outweighs financial motives for explaining socially responsible investment decisions based on a survey conducted in 2011 with Dutch investors. The impact of non-pecuniary factors is also supported by Hartzmark and Sussman (2019), who evaluated the effect of the first introduction of sustainability ratings by Morningstar for the U.S. mutual fund market in 2016, which supported the evaluation of a funds' sustainability. They found that fund flows for more sustainable

funds were positively affected, while the flows to less sustainable funds decreased.

In light of this, there is a growing branch of studies evaluating the size of a potential green premium. Reviewing 15 publications that have been published in this area between 2007 and 2019, MacAskill et al. (2021) report a lack of consensus regarding the existence of such premium, which they attribute to different methodological approaches. Nevertheless, its presence is reported in the majority of the studies for both, the primary market (56%) and the secondary market (70%). For the latter market, the reviewed studies observe an average green premium of -1 to -9 basis points (bps). Further, MacAskill et al. (2021) highlight that the premium is generally more profound for green bonds that are "government issued, investment grade, and that follow defined green bond governance and reporting procedures". For the latter, they provide recognized green bond certification standards and third-party verification for the use of proceeds as main drivers. Hachenberg and Schiereck (2018) argue that such an enhanced reporting is necessary to mitigate information asymmetries between the issuers and investors. This aligns with the results of a survey conducted by the CBI (2019) with 48 of the largest Europe-based fixed income asset managers, which showed that they view green credentials, next to pricing, as one of the most important factors for their investment decisions. This is also consistent with Dorfleitner et al. (2021), whose findings support the positive effect of external validation on the green premium. Moreover, Immel, Hachenberg, Kiesel, and Schiereck (2021) and Hachenberg and Schiereck (2018) find the issuer's Environmental, Social and Governance (ESG) rating to influence the yield differential between green and conventional bonds. Finally, Kapraun, Latino, Scheins, and Schlag (2019) identify a bond's "Green credibility" as a main driver for the green premium. The German Green Federal Securities, which are the focus of this dissertation, seem to fulfil the above-mentioned driving factors. However, the evidence for the existence of a green premium for sovereign green bonds considering both, the primary market and the secondary market, is not conclusive. For example, while Doronzo, Siracusa, and Antonelli (2021) find no definite evidence for such a premium based on bond data from 14 countries that have been issued between end-2016 and 2020, Kapraun et al. (2019) find a significant green premium between 5 and 18.5bp for bonds that are issued by government entities.

A feasible methodological approach to evaluate the premium of green bonds is the comparison with a counterfactual brown (i.e., non-green) bond that otherwise exhibits the same characteristics (Bachelet, Becchetti, & Manfredonia, 2019). As such a security is in general not available, one viable alternative is to find a proxy based on a matching method. For example, Bachelet et al. (2019) identify brown nearest-neighbours that have the same currency, issuer, rating coupon type and a similar maturity date, coupon rate and amount issued. Doronzo et al. (2021) and Zerbib (2019) also use a direct matching approach by constructing a synthetic brown bond based on other bonds that have similar properties. Alternatively, two-step matching procedures such as propensity score matching (e.g., Gianfrate and Peri, 2019) or coarsened exact matching (see Löffler, Petreski, and Stephan, 2021) are applied to obtain estimates for the "untreated" brown bonds. However, in this study it is not necessary to rely on proxies for a counterfactual brown bond, as the German Green Federal securities are issued as twins to conventional bonds that share most of their characteristics.

2.2. Liquidity Premium

One property that differs is that German Green bonds have a lower issuance volume than their conventional counterparts (Finanzagentur GmbH, 2021b). The Finance Agency (2021a) argues that a sufficiently high amount outstanding is necessary to ensure that they can be traded in large quantities and at any time. This is because a low volume can imply a lower liquidity due to less owners and thus higher search costs (Helwege, Huang, & Wang, 2014). Therefore, investors may require a higher return to compensate for the additional risk of holding an illiquid asset (Kempf & Uhrig-Homburg, 2000). This understanding of liquidity is based on Fisher (1959), who views it as the ability to sell a bond quickly and without a discount on its value.

The impact of illiquidity on bond prices in general is widely researched (e.g., Chen, Lesmond, and Wei (2007); Dick-Nielsen, Feldhütter, and Lando (2012); Helwege et al. (2014); Kempf and Uhrig-Homburg (2000); Schestag, Schuster, and Uhrig-Homburg (2016)). A main advantage of understanding and measuring illiquidity costs is that it can help investors to improve the management of their exposure to risks. For example, investors who hold a bond until maturity (i.e., no need to sell it early) are not affected by liquidity disadvantages and thus may favour a premium for holding an illiquid asset (Wegener, Basse, Sibbertsen, & Nguyen, 2019). However, if the premium is attributed to other factors (e.g., credit risk) instead, this may not be the optimal investment alternative for such investors.

Nevertheless, it is not straightforward to derive a feasible proxy for the liquidity, which can be attributed to the lack of a universal definition (Díaz & Escribano, 2020). Díaz and Escribano (2020) provide an overview on the various dimensions of liquidity and the selection of proxies that measure its different characteristics. In the context of this dissertation, a proxy that indicates the size of illiquidity costs over time is required. One viable approach to estimate this liquidity premium is the comparison of yields of bonds that only differ in their degree of liquidity. While Schwarz (2019), Monfort and Renne (2014) and Schuster and Uhrig-Homburg (2012) compare the liquid German Federal Securities with less liquid bonds from the German state-owned investment and development bank, Kreditanstalt für Wiederaufbau (KfW), Kempf, Korn, and Uhrig-Homburg (2012) compare the German Federal Securities with less liquid Pfandbrief bonds (Covered Bonds) and Wegener et al. (2019) compare less liquid traditional Pfandbrief bonds with Jumbo Pfandbrief bonds that have a larger issuance volume. To relate the liquidity premium to different investment horizons, Kempf et al. (2012)

model the premium based on the Nelson and Siegel (1987) approach, while Koziol and Sauerbier (2007) use the Svensson (1994) method. Both parametric models provide the term structure of the current spot rate for zero coupon bonds.

The impact of differences in liquidity is also considered in the context of green bonds. For example, Zerbib (2019) remarks the explanatory power of a liquidity proxy for the yield differential between green bonds and counterfactual conventional bonds, when estimating the green premium. Further, Bachelet et al. (2019) find evidence for a higher liquidity of green bonds issued by public institutions in relation to their brown (i.e., conventional) counterparts. Finally, Wulandari, Schäfer, Stephan, and Sun (2018) find a negligible impact of liquidity risk on green bonds.

Finally, liquidity risks can affect the financing costs for issuers of bonds. The costs of capital are determined by the primary market yields issuers can secure at issuance. However, the return investors require from holding a bond may be affected by its expected performance on the secondary market. For example, Goldstein, Hotchkiss, and Pedersen (2019) find evidence based on corporate bonds that the expected aftermarket liquidity at issuance can have an economically large impact on the financing costs. A viable explanation for this finding is that the initial investors have a lower perceived risk in case they need to sell the asset before its maturity, and are thus willing to pay a premium. From an issuer's perspective, it can therefore be advantageous to ensure a liquid secondary market for its bonds.

2.3. Term Structure Models

To derive a structural model for the effect of illiquidity on the German green bonds, we assume a stochastic model that can reflect the development of the liquidity premium until maturity. For this purpose, we apply a term structure model of the short rate, which provides their development over time. In other words, we use a stochastic process to model a sequence of interest rates (i.e., a liquidity premium) each for an infinitely small period of time. This type of models are widely applied to value interest rate derivatives, such as European bond options (Hull, 2018). Further, they have also been used in studies that model bond illiquidity (e.g., Kempf and Uhrig-Homburg (2000); Koziol and Sauerbier (2007)).

In general, the various approaches can be divided into equilibrium models and no-arbitrage models (Hull, 2018). The drift of the short rate in equilibrium models is no function of the time, whereas the drift in no-arbitrage models is time-dependent. While the latter approach allows an exact fit to the current term structure, this is not required for the present application (Hull, 2018). This is because the used liquidity proxy is not calibrated to the actual (il-)liquidity of German green bonds, which only allows a relative evaluation of the effects. Hull (2018) presents the Rendleman and Bartter (1980) model, the Vasicek (1977) model and the Cox, Ingersoll, and Ross (1985) model as possible equilibrium models. The Rendleman and Bartter model differs in a way from the other models that it does not assume a mean reversion for the short rate, while the Cox, Ingersoll, and Ross model excludes negative interest rates by construction. The Vasicek model assumes a mean-reverting process for the short rate and allows for negative rates.

3. Green Sovereign Bonds

The CBI reports that 22 national governments have issued sovereign Green, Social, and Sustainability (GSS) bonds until November 2020 with a total amount of 96 bn. USD (Harrison & Muething, 2021). In the same study that was published in January 2021, Harrison and Muething (2021) report for the majority of GSS bonds a relatively higher imbalance between their supply and demand compared to their conventional counterparts, which was suggested by a mostly higher book cover (i.e., oversubscription). This indicates market growth potential for the green sovereign bonds. Moreover, they report that based on 23 issuances between 2017 and November 2020, ten bonds priced on the yield curve of conventional peers, nine priced below and four above. As the green bond issuance at a yield below the yield curve of conventional (i.e., vanilla) bonds implies more favourable financing costs, the observed sovereign bonds provide no clear evidence for such a potential yield advantage. In the following, we first provide an insight into the German green bond framework. Then we compare it to a small peer group of sovereign green bonds with a focus on how potential liquidity issues are addressed.

3.1. German Green Federal Securities

Since September 2020, the German federal government issued Green federal securities with a total volume of 24 bn. EUR (see Table 1). In 2021, it issued 12.5 bn. EUR of Green bonds, which accounted for 2.6% of the total issuance volume (482.7 bn. EUR) of tradable government debt that year (Finanzagentur GmbH, 2021d). For 2022, it anticipates a similar annual volume (Finanzagentur GmbH, 2021a).

The German Finance Agency ("Finanzagentur"), which administers the issuance of green bonds in the primary market, acknowledges the need to account for sustainability in financial decisions in light of economic risks as well as investment opportunities that result from climate change and transition towards a "more sustainable global ecosystem" (Finanzagentur GmbH, 2020). It concludes that this requires an enhanced transparency and development of the market for green and sustainable investments, to which the Green Federal securities are a key driver.

On the one hand, the enhanced transparency can be attributed to the chosen evaluation, selection, and reporting process. The criteria to identify eligible budget items align with the Green Bond Principles by the ICMA, the EU Charter of Fundamental Rights and consider elements of the draft EU Green Bond Standard (European Commission, 2012; Finanzagentur GmbH, 2020; International Capital Market Association, 2021).

In its Green Bond Investor Presentation (2021b), the Finance Agency provides an overview of the use of proceeds of the German Green Bonds. For example, it attributes the eligible expenditures in 2019 to the following sectors: Transport

Table 1: Overview of German Twin Federal Securities

Name	Issuance	Maturity Date	Outstanding	Type	ID
2021 (2050) Bund/g	18.5.2021	15.8.2050	6.0 bn. EUR	Green	G2050
2019 (2050) Bund	23.8.2019	15.8.2050	29.0 bn. EUR	Conventional	C2050
2021 (2031) Bund/g	10.9.2021	15.8.2031	6.5 bn. EUR	Green	G2031
2021 (2031) II Bund	18.6.2021	15.8.2031	26.5 bn. EUR	Conventional	C2031
2020 (2030) Bund/g	9.9.2020	15.8.2030	6.5 bn. EUR	Green	G2030
2020 (2030) II Bund	19.6.2020	15.8.2030	30.5 bn. EUR	Conventional	C2030
Bobl/g	6.11.2020	10.10.2025	5.0 bn. EUR	Green	G2025
Bobl	10.7.2020	10.10.2025	25.0 bn. EUR	Conventional	C2025

The data in this table is based on Finanzagentur GmbH (2021a) and Refinitiv Eikon (Accessed: 21.11.2021). Table 16 in the Appendix shows an extended version of this table.

(57.9%), International cooperation (24.2%), Energy (9.7%), Research (5.1%) and Agriculture (3.1%). Moreover, it highlights amongst other eligible expenditures the upgrade and electrification of the railway between Ulm and Lindau with total costs of approx. 225 million (mn.) Euro (EUR) and the development loan for a renewable power plant (i.e., solar PV) in India amounting to 89.3 mn. EUR as examples from the infrastructure and international cooperation sector, respectively.

The evaluation criteria are only applied to expenditures that are already accrued (Finanzagentur GmbH, 2020). This process enhances the transparency of the use of proceeds, as the projects precede the issuance of the securities. However, it also restricts the issuance amount of green bonds as, for example pending expenditures are not eligible. In addition to the selection criteria, the agency provides a Second Party Opinion on the Green Bond Framework (see ISS ESG, 2020), an external Third Party Verification of the Allocation Report by the auditing firm Deloitte, and impact reporting (Finanzagentur GmbH, 2021c).

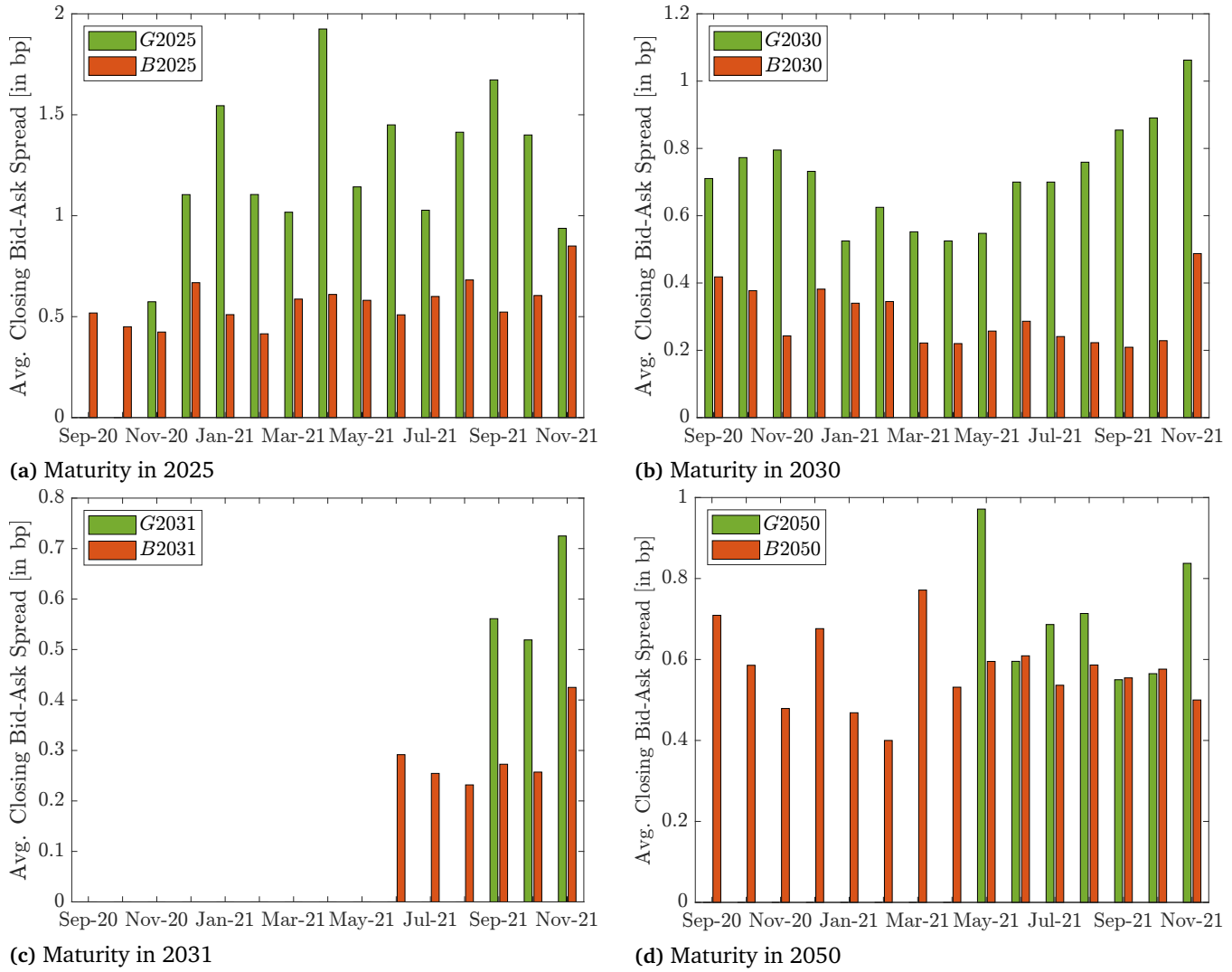
On the other hand, the German Green Bonds aim to support the development of the European green fixed income market by establishing a new interest rate benchmark for such assets, a green yield curve (Finanzagentur GmbH, 2020). While conventional government bonds with a high credit rating can be used to serve as a benchmark return for risk-free investments only, the green yield curve can provide the term structure of interest rates for riskless and green assets as they are ranked pari-passu (i.e., equally) to the conventional bonds from the same issuer. This means that they could provide a reference for the required future payoff of a risk-free green investment with a specific time horizon.

To provide this information to potential investors and quantify their preference for green investments, the German Green Bonds are issued based on a unique twin bond concept (Finanzagentur GmbH, 2020). As summarized in Table 16 in the Appendix, the green bonds and their conventional twins share the same coupon rate and time to maturity, but differ in their issuance volume and are traded separately (i.e., they have different ISIN codes).

In addition to the high credit quality, another requirement of German Federal securities to serve as benchmark is sufficient liquidity (Finanzagentur GmbH, 2021a). This is to mitigate risks for bondholders that can be induced by illiquidity, for example, the inability to sell the bond rapidly or only for a lower transaction price (e.g., see Kempf and Uhrig-Homburg, 2000). In the context of green bonds, their issuance can entail liquidity risks for both, themselves and conventional twins. This is because a sufficiently high amount outstanding of each type of bond is necessary to ensure that they can be traded in large quantities and at any time (Finanzagentur GmbH, 2021a). While a low issuance volume of green bonds may impede their own liquidity, a high volume can have detrimental effects on the liquidity of conventional bonds, if the total outstanding volume of Federal securities is maintained (Finanzagentur GmbH, 2021a). As a consequence, this potential trade-off has to be solved in order to provide an interest rate benchmark for both, the green as well as the conventional European fixed income market.

The figures in Table 1 indicate that the volume of each green bond is significantly smaller than its conventional counterpart. In fact, the average amount outstanding of a German green bond is approximately one fifth (22%) of the amount of the average conventional twin bond. This suggests that the green bonds may be less liquid than their conventional twins. To test this hypothesis, we evaluate the bid-ask spread (BAS) of the daily closing bond yields, as this measure is frequently used to derive a proxy for the liquidity of bonds (e.g., Dick-Nielsen et al. (2012); Kapraun et al. (2019); Zerbib (2019)). A higher BAS represents higher transaction costs, which can indicate a lower liquidity. Figure 1 shows that since the issuance of the first German green bond, the average monthly BAS of the green bonds is almost consistently larger than the spread of the conventional counterparts.

To verify this visual impression, a paired t-test is performed whether the average bid-ask spread ($\overline{BAS}$) since issuance of each green bond $\overline{BAS}_G$ coincides with the same measure for the respective conventional twin $\overline{BAS}_C$. The test results in Table 2 show that we can reject the null hypothesis



(a) Maturity in 2025

(b) Maturity in 2030

(c) Maturity in 2031

(d) Maturity in 2050

Figure 1: Average monthly closing bid-ask spreads for German twin bonds

The figure is based on data from Refinitiv Eikon (Accessed: 10.11.2021) and shows the average monthly closing bid-ask spread in basis points (bp) for German green bonds and their conventional twins displayed in Table 1 for the period from 09.09.2020 to 10.11.2021. Summary statistics are displayed in Table 11 in the Appendix.

H_0: $\overline{BAS}_G = \overline{BAS}_C$ on a significance level $\alpha = 0.05$ for all twins, in favour of the alternative hypothesis H_A: $\overline{BAS}_G \neq \overline{BAS}_C$. Therefore, we can conclude that the green bonds are traded during the observed period from 09.09.2020 to 10.11.2021 on average at a statistically significant wider spread than the conventional twin bonds.

The absolute difference between the average spreads Δ ranges from 0.1 to 0.7 basis points. To evaluate the economic significance of this difference, we assume values that align with the twin bonds that mature in the year 2031. Assuming a bid yield $y_{bid} = -40$bp for both twins, a narrower spread for the conventional bond with $y^C_{ask} = -39.7$bp and a wider spread for the green bond $y^G_{ask} = -39.3$bp, we have $\Delta = 0.4$bp representing the additional transaction costs TC. Further, we assume a time to maturity of $T = 10$ years. To compute the present value P of a zero-coupon bond, we assume continuous compounding to discount its nominal value N and thus use $P = N \cdot e^{-yT}$. We obtain the trading costs as $TC = N \cdot (P_{ask} - P_{bid})$. Based on this specification, an investor, which acts as a price taker and executes a round trip trade with both, a German green bond and its conventional twin, by buying at the ask price P_{ask} and selling at the bid price P_{bid} for an investment of $N = 1$ mn. EUR and $T = 10$ years, would incur additional trading costs for the green bond amounting to $\Delta TC = TC_G - TC_C = N \cdot (P^G_{ask} - P^C_{ask}) = 385$ EUR or about 4bp that are caused by its wider spread. From an economical perspective, this amount is relatively small, which aligns with the objective of the German approach to address liquidity risks. However, in relative terms, $\overline{BAS}_G$ is on average almost twice the size (+94%) of $\overline{BAS}_C$ for the observed data. It should be noted that these transaction costs are different to the liquidity premium in the model that de-

Table 2: Closing bid-ask spreads for German twin bonds

	BAS_{2050}	BAS_{2031}	BAS_{2030}	BAS_{2025}
$\overline{BAS}_C$ [in bp]	0.573	0.294	0.290	0.572
$\overline{BAS}_G$ [in bp]	0.669	0.570	0.703	1.273
$\Delta = \overline{BAS}_G - \overline{BAS}_C$	0.096	0.276	0.413	0.701
t-statistic	2.9223	5.5221	17.8084	12.4163
p-value	0.0041	0.0000	0.000	0.0000
N	131	47	305	262

The table shows the results of a paired t-test to determine whether the mean bid-ask spread for the closing yields of German green bonds $\overline{BAS}_G$ coincides with the same measure for the conventional twins $\overline{BAS}_C$. The null hypothesis H_0: $\overline{BAS}_G = \overline{BAS}_C$ is tested against the H_A: $\overline{BAS}_G \neq \overline{BAS}_C$. The data is retrieved from Refinitiv Eikon (Accessed: 10.11.2021) and covers the period from 09.09.2020 to 10.11.2021. Summary statistics are displayed in Table 11 in the Appendix.

scribes a premium to the yield (i.e., a higher yield) for illiquid assets instead.

The German green bond concept is designed to mitigate possible liquidity disadvantages for both twins. To ensure the liquidity of the conventional bonds, the Finance Agency issues the same amount as the green counterparts in its own stock, which can be used on the secondary market for repurchase agreements (i.e., repo transactions) and lending activities (Finanzagentur GmbH, 2021a). Therefore, the total amount of conventional securities and thus their liquidity remains unchanged. To ensure the liquidity of the green bonds, the German Finance Agency declares to engage in secondary market activities (Finanzagentur GmbH, 2020). In their Investor Presentation from September 2021, the Finance Agency categorises them as (1) Outright ("one-way") sales and purchases, (2) Repurchase agreements and securities lending, using the Federal Government's own stock of green bonds and (3) Combined and debt-neutral sale-and-purchase (switch) transactions conducted with banks that are members of the Bund Issues Auction Group (Finanzagentur GmbH, 2020, 2021b). This means that it can influence the supply and thus the price of the green bonds on the secondary market. From the issuer's perspective, green bonds are more valuable than the conventional twins. Although both zero-coupon bonds have the same face value and thus the same cash flows, this is because the green bonds provide an additional documentation for the usage of their proceeds. Even in situations, where investors would not attribute a higher value to the green bonds, this would still hold for the issuer, who sustains the associated added costs and more limited use of proceeds. Therefore, the switch, which is the simultaneous and debt-neutral sale of a conventional bond and purchase of a green bond, would be economically viable for the issuer at a yield spread of zero. Further, it can execute such transactions until the green securities are completely in their own holdings. In this case, the respective amount of conventional twins that was initially held back by the Finance Agency is then traded in the secondary market.

Figure 2 shows all available closing ask yields until 04.11.2021 of the German twin bonds displayed in Table 1 retrieved via Refinitiv Eikon. It shows that the yields of the respective twins are closely related for the observed period of time. Further, the data supports an upward sloping yield curve for both bond types. This means that investors with a longer time to maturity require a higher rate of return, ceteris paribus. Figure 3 shows the yield differential between German green bonds y_G and their respective conventional twin y_C (i.e., $\Delta y = y_G - y_C$). For most of the observed period, the data shows a negative trending spread with an average of around -5bp. This implies that investors are increasingly willing to sacrifice return in favour of investing into the German Green Federal securities. However, it should be noted that the historical data covering a period of one year is relatively scarce and the future size of the spread may change.

3.2. Addressing Liquidity Risks

This section aims to provide a brief insight into how selected sovereign issuers different to Germany address the possible risk of illiquidity. These issuers are France, the Netherlands, and Belgium. France issued its first green sovereign bond in 2017 for 7 bn. EUR, which was since then increased to a total amount outstanding of 28.9 bn. EUR République Française (2021). In the French framework for green Obligations assimilables du Trésor (OAT) (2017), their liquidity is emphasized on its first page. Also, the respective investor presentation covers the liquidity as one of six main topics (République Française, 2021). This underlines the relevance of liquidity concerns. In the same document, they show that the average monthly bid-ask spread is consistently lower for their green bond (RIC: FR0013234333=) than a conventional bond (RIC: FR0013515806=) which matures one year later in 2040. Further, both bonds show a similar ownership structure with a share of long-term investors of 37% and 38%, respectively. They also highlight that its amount outstanding is with 31 bn. EUR similar to neighbouring (in terms of time to maturity) conventional bonds and argue that this supports its liquidity (République Française,

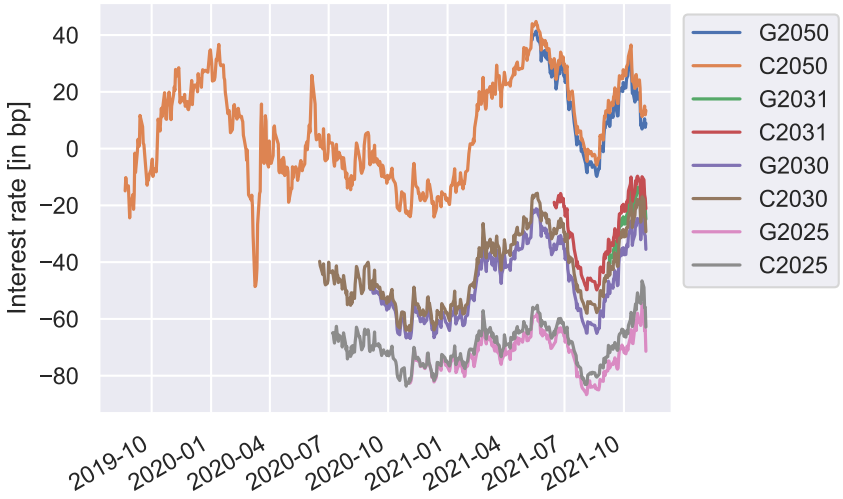


Figure 2: Yield of German government bonds and green twins

The figure is based on data from Refinitiv Eikon (Accessed: 04.11.2021) and shows the closing ask yield in basis points for the German Federal securities displayed in Table 1. The summary statistics are displayed in Table 12 in the Appendix.

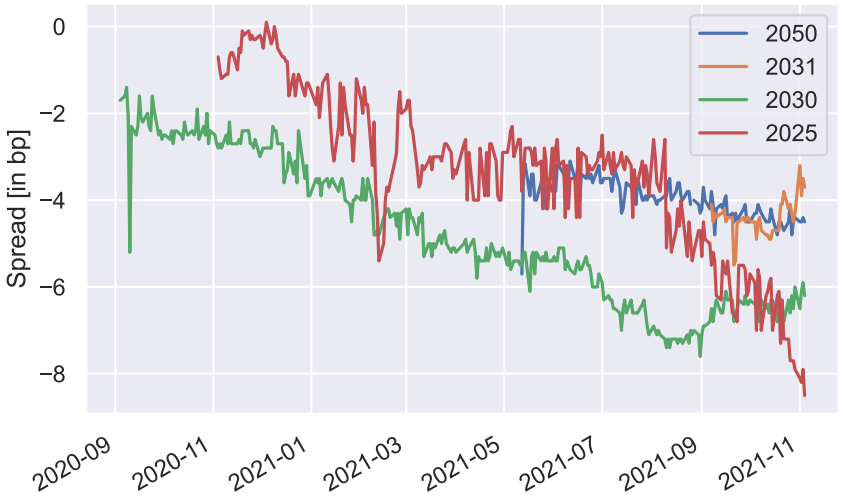


Figure 3: Spread between German government bonds and green twins

The figure is based on data from Refinitiv Eikon (Accessed: 04.11.2021) and shows the yield differential of German green bonds y_G and their respective conventional twin y_C (i.e., $y_G - y_C$) in basis points. The summary statistics are displayed in Table 13 in the Appendix.

2017, 2021). The Dutch State Treasury Agency (2019) justifies the liquidity of the Dutch sovereign green bonds with a minimum issuance volume of 10 bn. EUR within several years, a quotation obligation for Primary Dealers to ensure the availability of tradable prices and a Repo facility available to Primary Dealers that serves as a lender of last resort. Belgium reports that its green bonds have no liquidity disadvantages and a similar issuance volume as conventional government bonds (The Kingdom of Belgium, 2018).

So far, there exists only one other country that decided to adopt the German approach that was introduced in 2020. The national bank of Denmark 2022 reported following the

German twin bond concept with the first Danish green government bond that was issued on January, 19th 2022 as a 10-year zero coupon bond. Further, they announce that a switch of the green bond to its corresponding more liquid conventional twin bond will be possible for investors "at any time" to support its liquidity. Before opting for the twin bond concept, the Danish Debt Management Office also considered issuing green certificates in addition to the conventional bonds instead (Bongaerts & Schoenmaker, 2020). Bongaerts and Schoenmaker (2020) recommend such green certificates as a viable approach to meet the demand for environmentally-friendly debt, while avoiding potential drawbacks of green government bonds. Namely, impeding the liquidity of both, green and conventional bonds, making the price of green certificates more suitable to adequately reflect environmental fundamentals.

The above examples suggest that possible liquidity concerns are a relevant factor, which is generally addressed by issuers of green bonds. The issuing institutions of French, Dutch and Belgian sovereign green bonds all emphasize a sufficiently high amount outstanding as one mitigating measure of liquidity disadvantages. However, a large amount of green bonds can potentially have adverse effects on the liquidity of conventional bonds and might thus not be desired (e.g., see Finanzagentur GmbH, 2021a). An evaluation of the potential post-issuance liquidity effects of green bonds on conventional bonds from those countries is not pursued in this dissertation due to its limited scope. In the case of Germany, the largest currently traded German green bond, with an issuance size of 6.5 bn. EUR, is relatively small compared to the Dutch benchmark of 10 bn. EUR. Therefore, it is reasonable that the German twin bond approach aims to offer an alternative approach to address such risks.

4. Methodology and Data

In this section, we derive a non-closed form solution for the yield differential (i.e., spread) between German green bonds y_G and its conventional twin bonds y_C. To achieve this, we decompose the yield into three effects of the green bond relative to the conventional bond: A liquidity premium LP, a green premium GP, and the effect of the secondary market interventions (i.e., switch transactions) by the German Finance Agency, in the following denoted as ST. Therefore, we write the decomposition of the yield differential Δy as

$$\begin{aligned}\Delta y &= y_G - y_C \\ &= LP - GP - ST.\end{aligned} \tag{1}$$

For any additional degree of illiquidity of the green bond, investors require a higher return (i.e., a higher LP), which increases the spread. Further, investors may accept a lower yield for investing into a “green” asset (i.e., a higher GP), which reduces the spread. Finally, the market interventions (i.e., switch transactions ST) by the Finance Agency increase the liquidity of the green bonds and thus have a negative effect on the spread as well. It should be noted that a negative green premium (e.g., as in Zerbib (2019)) is in Equation 1 defined as a positive value for GP and thus subtracted. In the same fashion as Kempf and Uhrig-Homburg (2000), this model assumes perfect and arbitrage-free markets except for illiquidity costs. However, the bonds are traded in discrete time and only the liquidity premium is subject to change, which is modelled as a stochastic short rate. The green premium and the interest rate of the conventional bond are assumed to be constant.

4.1. Trinomial Tree Model

We consider a stochastic liquidity premium LP_t that follows an Ornstein-Uhlenbeck process, as suggested by Vasicek (1977). Based on this approach, the increment dLP_t is defined as

$$dLP_t = a(b - LP_t)dt + \sigma dz, \tag{2}$$

where a, b are non-negative constants and denote the mean reversion rate and the long term level reversion level, respectively. σ denotes the local volatility and dz follows a standard Wiener process with $dz = \varepsilon \cdot \sqrt{dt}$ and $\varepsilon \sim \mathcal{N}(0,1)$. From a theoretical perspective, it is plausible to assume a mean-reverting process for the liquidity premium, as a lower level of liquidity leads to a higher premium, which may attract new investors. This increase in demand for the bond can positively affect its traded volume on the secondary market and thus increase its liquidity. A geometric Brownian motion would not coincide with this theoretic argumentation, as the liquidity premium could increase (or decrease) indefinitely. Nevertheless, a Dickey-Fuller test is performed during the calibration of the model to confirm if the discrete data for the selected liquidity proxy supports a random walk or not.

In the next step, we derive a discrete trinomial tree representation of the stochastic process $\tilde{LP}$. This non-closed form solution is required to incorporate the effects of ST into the model for the green bond yields. We can re-write the liquidity premium as $\tilde{LP} = b + \tilde{s}$, where b denotes the long term mean as in Equation 2 and $\tilde{s}$ denotes the stochastic part of the premium. Using this, we can rewrite Equation 2 as

$$\begin{aligned}dLP_t &= a\,(b - (b + s_t))\,dt + \sigma dz \\ &= -a s_t dt + \sigma dz,\end{aligned} \tag{3}$$

with $\mathbb{E}[dLP_t] = -a s_t dt$ and $Var[dLP_t] = \sigma^2 dt$. We use this result to derive the trinomial tree representation, where the change in $\tilde{s}$ for each time step is indicated by Figure 4.

The spacing between the nodes in the time-dimension i (e.g., $s_{j,i}$ and $s_{j,i+1}$) is denoted as $\Delta t = \frac{T}{N}$, for a tree with an investment period of T years and N equidistant discrete time steps. The spacing between the state-dimension j with $s_{j+1,i} - s_{j,i} = s_{j,i} - s_{j-1,i}$ is denoted as Δs. This means that node (j,i) describes the possible states that the liquidity premium can assume in time $t = i\Delta t$ with $LP_{j,t} = b + s_t = b + j\Delta s$. The probabilities for the up-state, mid-state and down-state in the next period are denoted as p_u , p_m & p_d, respectively.

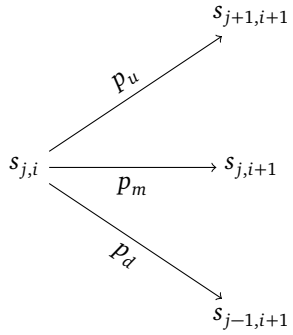


Figure 4: Trinomial tree of stochastic part of liquidity premium $\tilde{s}$

Following Hull (2018), we set the spacing Δs to

$$\Delta s = \sigma\sqrt{3\Delta t}, \tag{4}$$

which was found to minimize the model error. Further, we restrict the branching structure to ensure positive probabilities in the tree (Hull, 2018). For this, we set the limits j_{max} and $j_{min} = -j_{max}$, where the branching changes from the form displayed in Figure 4 to the respective structure displayed in Figure 5. In the same fashion as Hull (2018), we set j_{max} as the smallest larger integer than $\frac{0.184}{a\Delta t}$. While these limits restrict the maximum and minimum size of the mean-reverting liquidity premium, they do not impede the fitting of the discrete trinomial tree model to the continuous-time Vasicek process from Equation 2. The calibrated tree is still able to match the first two moments of the observed process.

We derive the discrete solution for the liquidity premium by setting three restrictions on the three time-independent, but state-dependent tree probabilities p_u, p_m and p_d (Hull, 2018). In detail, for each time-step, we match the first two moments of ds_t, using the expected change $\mathbb{E}[dLP_t] = -as_t dt$ and the variance $Var[dLP_t] = \sigma^2 dt$. Further, we require the probabilities to add to one.

For the default branching method (i.e., $j_{min} < j < j_{max}$) the condition for the expected change notates as,

$$\begin{aligned} p_u \cdot \Delta s + p_m \cdot 0 + p_d \cdot (-\Delta s) &= \mathbb{E}[dLP_t] \\ &= -a \cdot j \cdot \Delta s \cdot \Delta t \end{aligned} \tag{5}$$

For the condition for the variance we use $Var[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$ and thus obtain

$$\begin{aligned} \mathbb{E}[dLP_t^2] &= Var[dLP_t] + \mathbb{E}[dLP_t]^2 \\ p_u \cdot \Delta s^2 + p_m \cdot 0^2 + p_d \cdot \Delta s^2 &= \sigma^2 \cdot \Delta t + a^2 \cdot j^2 \cdot \Delta s^2 \cdot \Delta t^2 \end{aligned} \tag{6}$$

The final condition is for all branching structures the same and denotes as

$$p_u + p_m + p_d = 1. \tag{7}$$

The expressions for the probabilities for each branching structure are derived in Appendix **??** and coincide with the solution provided by Hull (2018).

4.2. Extension to Twin Bond Approach

In Section 4.1, we derived a discrete trinomial tree model that provides the respective probability weights for the change of the stochastic liquidity premium at each time step. In the following, we use this result to derive a solution for the initial bond price at time $t = 0$. For this, we assume no default risk. This implies that the price of each bond at time $t = T$ is set equal to its nominal value $P_T = 1$. To obtain the fair value of the bond in the periods before, we need to discount the expected bond price with the correct discount rate.

For example, for a liquid and non-green zero coupon bond C with a constant interest rate r we obtain its value at time t as,

$$P_t^C = \frac{P_{t+1}^C}{e^{r\Delta t}}. \tag{8}$$

This implies a present value at time $t = 0$ of $P_0^C = P_T^C \cdot e^{-rT} = e^{-rT}$, where $T = N\Delta t$.

This expression for the bond price changes considering the stochastic liquidity premium in addition to the interest rate r. In this case, the trinomial tree allows the derivation of state-dependent results based on the respective value of the premium. As before, we assume that the value of the bond must equal its nominal value at maturity in all states j of the liquidity premium. Therefore, the expected bond price at time $t = T$ is set equal to

$$\mathbb{E}[P_T|j] = P_T = 1. \tag{9}$$

To obtain the fair bond price at time t, we first need to compute its expected bond price at time $t + 1$ and then discount it with the correct interest rate. Following this approach, we can recursively obtain all bond prices until time $t = 0$. To compute the expected bond price we use the derived tree probabilities of the liquidity premium. Therefore, we can write the expected bond price at time $t = i\Delta t$ and state j, with $j_{min} < j < j_{max}$, in the trinomial tree as,

$$\mathbb{E}[P_{t+1}|j] = p_{u,j} \cdot P_{j+1,t+1} + p_{m,j} \cdot P_{j,t+1} + p_{d,j} \cdot P_{j-1,t+1}. \tag{10}$$

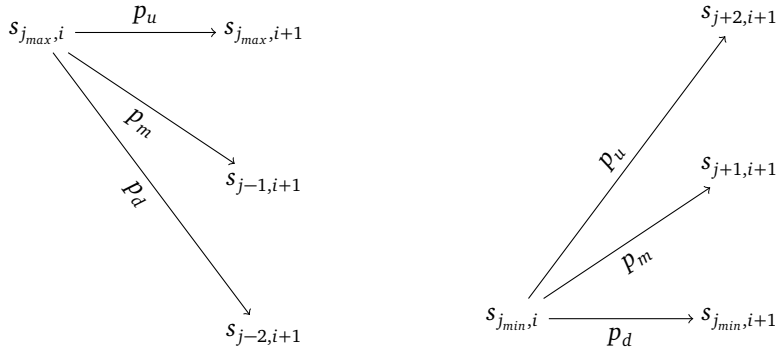


Figure 5: Restriction of trinomial tree branching structure

The figure is based on Hull (2018) and shows the upper and lower limits of the branching structer for $j = j_{max}$ (left) and $j = j_{min}$ (right), respectively.

For $j = j_{max}$ this equation changes to

$$\mathbb{E}[P_{t+1}|j_{max}] = p_{u,j_{max}} \cdot P_{j_{max},t+1} + p_{m,j_{max}} \cdot P_{j_{max}-1,t+1} + p_{d,j_{max}} \cdot P_{j_{max}-2,t+1} \tag{11}$$

and for $j = j_{min}$ we use

$$\mathbb{E}[P_{t+1}|j_{min}] = p_{u,j_{min}} \cdot P_{j_{min}+2,t+1} + p_{m,j_{min}} \cdot P_{j_{min}+1,t+1} + p_{d,j_{min}} \cdot P_{j_{min},t+1}. \tag{12}$$

For a counterfactual bond that only differs from the bond C in its liquidity, we need to consider in addition to r the respective liquidity spread $LP_{j,t}$. We obtain

$$P^I_{j,t} = \frac{\mathbb{E}[P^I_{t+1}|j]}{e^{(r+LP_{j,t})\Delta t}} \tag{13}$$

and use this expression to obtain the bond prices for all states and time steps in the tree via Backward-Induction. This is possible because we have the final value of the bond $P^I_T = 1$. By applying this procedure, we obtain one single value for the bond at time $t = 0$.

Based on this result, we can easily modify the expression in Equation 13 to additionally account for a constant green premium. This additional assumption implies for the model that the expected value that investors attribute to investing into a green asset does not change over time. We obtain

$$P^{IG}_{j,t} = \frac{\mathbb{E}[P^{IG}_{t+1}|j]}{e^{(r+LP_{j,t}-GP)\Delta t}}. \tag{14}$$

The negative sign implies that given a non-negative green premium GP, investors accept a lower yield to maturity. Discounting with a smaller value yields a higher price for the green and illiquid bond, which is therefore inversely related to the interest rate.

However, the expression for P^{IG} does not coincide with the theoretical value of the German green bond P^G. This is because it neglects the impact of the secondary market interventions by the FA. Section 3.1 outlines why it is not only possible but also rational for the FA to perform secondary market transaction, when the yield spread (i.e., $\Delta y = y_G - y_C$) between the twins assumes a non-negative value. In short, from the point of view of the FA, a green twin is due to the additional documentation for its use of proceeds always more valuable than the corresponding conventional twin. Therefore, if the yield of a green bond y_G notates above the yield of the conventional twin y_c, it is economically viable for them to execute combined and debt-neutral sale-and-purchase transactions. Those have a positive impact on the price of the green bond by reducing its supply, and thus negatively affect its yield and yield spread. We therefore assume that the price of a green bond cannot notate for a prolonged period of time below the price of its conventional twin. We can account for this additional characteristic by adding another condition to the model. Namely, we can restrict the prices of the German green bond to always assume values equal or higher than the corresponding conventional twin P^B_t. For the price at time t and in state j, this denotes as

$$P^G_{j,t} = \max\left[P^G_{j,t}, P^C_t\right]. \tag{15}$$

Further, this additional feature of the German green bonds can be interpreted as a call option on the illiquidity of the bond, assuming a constant green premium. When the liquidity premium becomes large, ceteris paribus, the switch transactions prevent P^G to fall below P^C. In this case, the value of the switch transactions need to compensate the lower price that would be implied by an illiquid and green bond alone. Therefore, above a certain threshold, increasing illiquidity leads to a higher value of the switch transactions. If the liquidity premium is sufficiently smaller than the green premium, intervention by the Finance Agency is unlikely to be required, and its value is equal to zero. Based on this comparison, we use in the following the terms "switch option" and "switch transactions" interchangeably to refer to the same mechanism of the German green bonds.

Building on the above results, we can use the model to derive the initial bond prices for a conventional bond, a coun-

terfactual bond with a liquidity premium, a counterfactual with a liquidity premium and a green premium and for the German green bond by accounting for the switch option. In general, we can define the initial by the model implied bond price as $P_{j=0,t=0} = P_0$ with

$$P_0 = P_N \cdot e^{-y_0 T}, \tag{16}$$

which we can reformulate to

$$\begin{aligned} y_0 &= \frac{\ln\left(\frac{P_N}{P_0}\right)}{T} \\ &= \frac{\ln\left(\frac{1}{P_0}\right)}{T}. \end{aligned} \tag{17}$$

Further, we can derive of the implied size for the liquidity premium LP, the green premium GP and the value of the switch transactions ST. We obtain the liquidity premium by subtracting the model yield of the conventional bond from the yield of the illiquid bond, which denotes as $LP = y_0^I - y_0^C$. Further, we obtain the value of the green premium by subtracting y_0^I from the yield of the illiquid and green bond, denoted as $GP = y_0^{IG} - y_0^I$. Finally, we can compute the value of the switch transactions by subtracting y_0^{IG} from the yield of the German green bond, denoted as $ST = y_0^G - y_0^{IG}$. As r is constant and has the same value for all bond types, it does not affect the implied values in the decomposition.

4.3. Model Calibration

In the following section, we calibrate the model parameters. For this, we need to estimate the parameters of the Vasicek model for the stochastic liquidity premium $\tilde{LP}$ and find a viable value for the green premium GP. The objective of this work is to evaluate the impact of secondary market interventions on the price formation of the green bond. For this purpose, it is not required to use exact estimates for the liquidity and green premium, but to focus on the size difference between both effects. This is sufficient because there exist infinite many combinations that yield the same result for the spread Δy. This can be shown by adding a constant m to the liquidity premium LP, and to the green premium GP to Equation 1, which cancel each other out.

To obtain an estimate for the development of the liquidity premium over time, we compare the yield differential of German Bundesanleihen (i.e., conventional bonds) and German Pfandbriefe (i.e., covered bonds). This is possible because they exhibit the same characteristics, but only differ in terms of liquidity. To adjust the results to the time horizons of the respective green bonds, we follow the approach suggested by Svensson (1994) to obtain estimates for the daily spot rate for an investment over T years, $y_{t,T}$. The required (daily) parameters are estimated and published by the Deutsche Bundesbank (2021) for both, conventional bonds and covered bonds. Following this method, the yield to maturity y_T at time t can be estimated by

$$\begin{aligned} y_T = \beta_0 &+ \beta_1\left(\frac{1-e^{-\frac{T}{\tau_1}}}{\frac{1}{\tau_1}}\right) + \beta_2\left(\frac{1-e^{-\frac{T}{\tau_1}}}{\frac{1}{\tau_1}} - e^{-\frac{T}{\tau_1}}\right) \\ &+ \beta_3\left(\frac{1-e^{-\frac{T}{\tau_2}}}{\frac{1}{\tau_2}} - e^{-\frac{T}{\tau_2}}\right), \end{aligned} \tag{18}$$

where β_0, β_1, β_2, β_3, τ_1 and τ_2 denote the daily estimated and published parameters by the Deutsche Bundesbank. Based on Equation 18, we obtain the daily estimates for the liquidity premium LP_t as,

$$LP_t = y_{t,T}^{Covered} - y_{t,T}^{Conventional}, \tag{19}$$

where $y_{t,T}^{Covered}$ denotes the estimated spot rate based on the daily parameters for the German Pfandbriefe, while $y_{t,T}^{Conventional}$ denotes the same measure for the German Bundesanleihen.

The Svensson method provides a daily measure for the historical development of the yield differential between German Pfandbriefe and Bundesanleihen for time to maturity T and serves as a proxy for the liquidity premium in the model. This assumption implies that the process for the long-rate coincides with the short-rate process of the liquidity premium that is modelled in Equation 2. This causes an estimation error, as the instantaneous liquidity premium can differentiate from the premium of longer maturities. For example, Kempf and Uhrig-Homburg (2000) found a higher liquidity premium in the longer-maturity segment. Further, this might also affect the mean-reversion and volatility characteristics of the assumed stochastic process. In terms of the model, a possible overestimation of the size of the premium does not affect the evaluation of the switch option. This is because the absolute difference between the GP and LP determines its value, which are evaluated for a range of spreads. Moreover, a sensitivity analysis of the model results to changes in the parameters is performed. To derive a viable proxy for LP, it requires two counterfactual interest rates that show a different liquidity premium. Other approaches to obtain such instantaneous proxy are to use shorter maturity times for the Svensson approach or follow Kempf and Uhrig-Homburg (2000) and compare the yield differences of two bonds that only differ in its liquidity and mature within the next year. In contrast to these measures, the chosen approach provides a liquidity proxy with the time horizon of the German green bonds. This has the advantage that it would be possible to remove the liquidity effects from the observed yield spread, by matching its size to the German green bonds.

Figure 6 shows the yield differential between German covered bonds and German conventional bonds for the same maturity times as the German green bonds, which serves as a proxy for the liquidity premium. The data aligns with the finding of Kempf and Uhrig-Homburg (2000) that the premium is larger for longer times to maturity. Further, the premium increases in 2Q2020 for the short to medium term bonds (i.e., 2025, 2030 & 2031), which may be attributed to

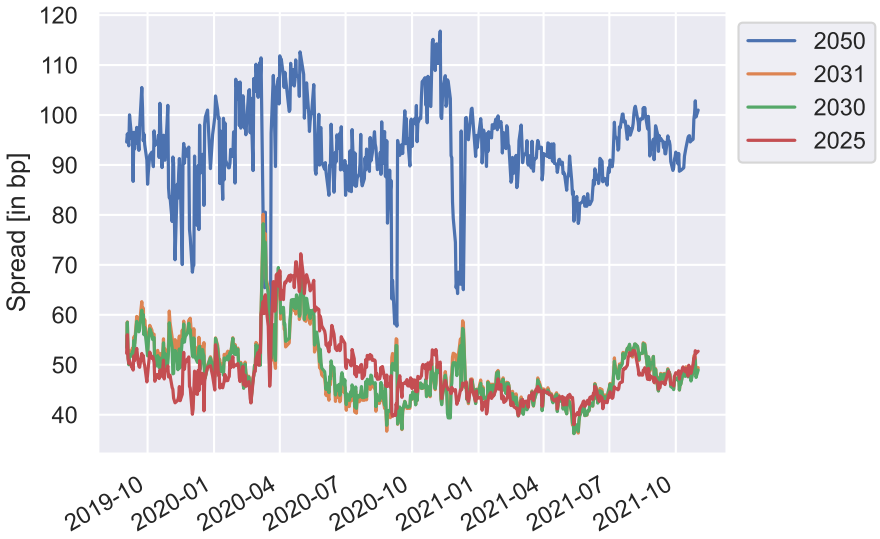


Figure 6: Yield spread between German government bonds and covered bonds

The figure displays the yield differential between German Pfandbriefe and Bundesanleihen. It is based on the estimated yield curves following the Svensson method and uses parameters published by Deutsche Bundesbank (2021). The summary statistics are displayed in Table 14 in the Appendix.

the effects of the Covid-19 crisis. A higher degree of uncertainty in this period could have increased the preference of investors to hold liquid assets, and thus the liquidity costs. While the volatility clusters in the data cannot be explained by the Ornstein-Uhlenbeck process in Equation 2 as it has a constant volatility term, the visualization in Figure 6 supports a mean-reverting process.

To test the adequacy of the underlying process for the given data, a Dickey-Fuller test is performed. This test can help to decide whether the data-generating process is stationary or has a unit root. As shown by Hayashi (2000), we need an ergodic stationary process to derive consistent parameter estimates for the population parameters. This is because a historical time series is only one possible realization of the underlying process. To obtain consistent estimates from the sample moments, we need to assume that all single observations over time result from the same process (i.e., stationarity) and that the memory of the process is not too persistent (i.e., ergodicity) (Hayashi, 2000). Further, a stationary process is also suggested by theory (see Section 4.1). Following Hamilton (1994), we estimate a random walk with drift and time trend,

$$\begin{aligned} \Delta LP_t &= LP_t - LP_{t-1} \\ &= \alpha + \gamma LP_{t-1} + \delta t + u_t, \end{aligned} \tag{20}$$

where α denotes the constant for the drift, γ is the coefficient for the unit root, δ denotes the slope of a linear time trend and u_t denotes independent white noise (i.e., an independent and identically distributed zero-mean error term with constant variance). To evaluate whether the data supports a unit root process, and thus does not support a mean-reverting process, we test the null hypothesis for the unit root $H_0 : \gamma = 0$ against the alternative hypothesis $H_A : \gamma < 0$. The relevant value of the test statistic τ is computed as,

$$\tau = \frac{\hat{\gamma}}{s.e.(\hat{\gamma})}, \tag{21}$$

which follows a non-standard distribution under the H_0. Therefore, we use the simulated critical values provided by Fuller (2009). The estimation results of the Dickey-Fuller test for the unit root parameter γ are summarized in Table 3. Based on the data, we can reject the H_0 of a unit root on a significance level $\alpha = 0.05$ for all time series, but for LP_{2025}. As a non-rejection of the H_0 contradicts the assumption for the data-generating process of the liquidity premium in the model, the results support the assumption in three of the four samples. We therefore proceed with the estimation of the model parameters specified in Equation 2.

We estimate the model parameters of Equation 2 to coincide with the Maximum Likelihood solution. This means that the estimated parameters b, a and σ maximize the joint probability that the estimated process yields the observed sample. For this, we follow Brigo, Dalessandro, Neugebauer, and Triki (2009) and estimate the parameters of the explicit solution for Equation 2 in discrete time by an Ordinary Least Squares (OLS) estimation,

$$LP_t = c + \phi LP_{t-1} + \delta\varepsilon_t, \tag{22}$$

where ε denotes Gaussian white noise. The estimation results of Equation 22 are summarized in the Appendix in Table 15. As suggested by Brigo et al. (2009), we use the

Table 3: Results of Dickey-Fuller Test

	LP_{2050}	LP_{2031}	LP_{2030}	LP_{2025}
$\hat{\gamma}$	−0.230	−0.136	−0.125	−0.047
$s.e.(\hat{\gamma})$	0.0304	0.022	0.0216	0.016
τ	−7.565	−6.114	−5.773	−2.935
$\tau^{crit}_{0.05}$	−3.423	−3.423	−3.423	−3.423
p-value	0.000	0.000	0.000	0.151
N	432	432	432	432

The table shows a summary of the Dickey-Fuller test results as specified in Equation 20. Observations with gaps due to missing data (i.e. weekends) are omitted.

following solution to obtain the parameters for Equation 2,

$$\begin{aligned} a &= -\frac{\ln(\phi)}{\Delta t} \\ b &= \frac{c}{1-\phi} \\ \sigma &= \frac{\delta}{\sqrt{(\phi^2-1)\Delta t/2\ln(\phi)}}, \end{aligned} \tag{23}$$

where we use $\Delta t = \frac{T}{N} = \frac{1}{250}$ due to daily observations. Further, we use a sample size of $N = 548$, which is lower than the sample of Kempf and Uhrig-Homburg (2000) that estimated the parameters of a term structure model using $N = 755$ observations. The sample period is chosen because it covers the complete period since the first emission of a German green bond. While a larger sample size might allow for a higher estimation precision, historical data that is too far in the past might not reflect current market conditions. The estimation results are summarized in Table 4.

The results in Table 4 for the fitted Vasicek process of the liquidity premium LP show significant differences based on their maturity time. The values for the mean reversion a and volatility parameter σ are larger for the curve that represents the long-term segment of the yield curve, namely 2050. This might be caused by the three drops shown in Figure 6 that are less pronounced and inversely seen for the short- and medium-term segments of the yield curves (i.e., 2025, 2030 & 2031). This means that during these very short periods, short-term liquidity became more expensive, while the long-term liquidity premium briefly declined in value but then returned to its initial level. As the model cannot accommodate such jumps and is fully described by the first two moments of the stochastic process $\tilde{LP}$, we rely on the estimated parameters for the 2025 time series for the further evaluation of the model.

Finally, the model requires a value for the green premium GP as an additional input parameter. In section 4.3 it was shown that the absolute difference between the long term mean of the liquidity premium b, in the following interpreted as the expected liquidity premium LP, and GP is sufficient to derive the size of ST, which is implied by the model. Therefore, for the purpose of evaluating the value of the switch option ST, it is only necessary to set the absolute difference between both premiums. For example, to reflect current market situations, we can set this difference so that the resulting yield spread coincides with the observed yield spread on the secondary market. Further, we can evaluate how the value of the switch option changes for different values of this spread between b and GP. For evaluating the effect of the switch option, it is thus not necessary to know the absolute value of the long term mean of the liquidity premium b, nor the value of the green premium GP, but only the difference $\Delta = LP - GP$. Similarly, the constant interest rate r affects both, the conventional bond C and the German green bond G, in the same fashion. Therefore, for this evaluation, an arbitrary value can be assumed as well.

5. Model Results

The evaluation of the model shows that it is able to reflect the main characteristic of the German twin bond approach. Namely, that the price of a German green bond P_G cannot fall below the value of its conventional twin P_C. Consequently, its maximum yield y_G is capped by an upper threshold equal to the yield of the conventional twin y_C. Furthermore, the model indicates additional potential advantages of the twin bond approach. Due to the additional value of the switch option, investors tolerate a higher degree of illiquidity until the value of the green bond assumes the threshold value of its conventional twin. In the same fashion, given a fixed level of illiquidity, a lower green premium is required that issuers can achieve a yield advantage, compared to a green and illiquid bond without a switch option.

5.1. Green Bond Yields

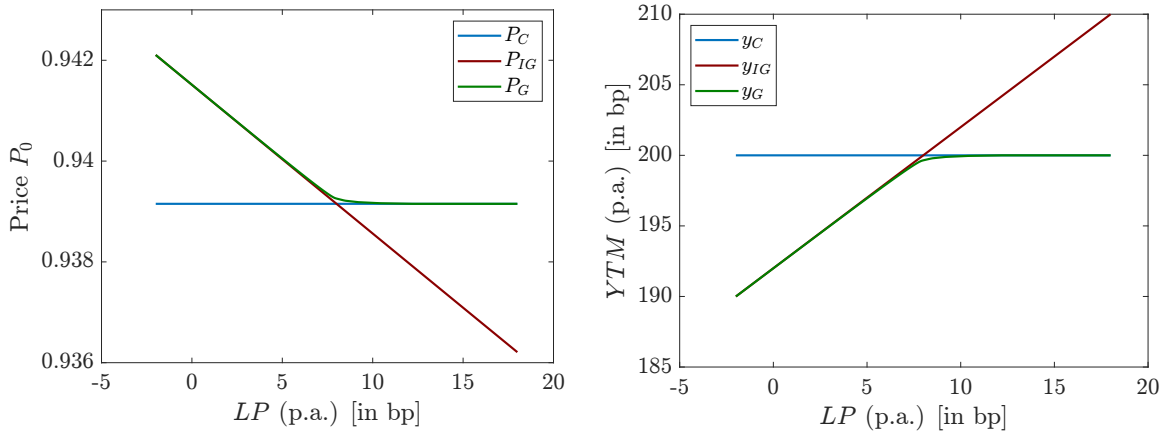
Using the parameters from the model calibration in section 4.3, we can evaluate the impact of changes in the expected liquidity premium LP for the different bond types. These are a conventional bond C, an illiquid and green bond IG, as well as a German green bond G, which is, in addition to being green and illiquid, also affected by the switch option.

Figure 7 shows the initial model bond prices and yields at $t = 0$, which are inversely related. Given a fixed face value

Table 4: Summary of ML estimates for $\tilde{LP}$

Vasicek model parameters		LP_{2050}	LP_{2031}	LP_{2030}	LP_{2025}
Mean-reversion rate	a	66.588	25.693	24.176	11.919
Long-term mean	b [in bp]	93.1	48.7	48.5	48.7
Instantaneous volatility	σ	0.0111	0.0046	0.0044	0.0031
Sample size	N	548	548	548	548

The table shows the estimation results for the process of the liquidity premium. The data is based on published yield curves by the Deutsche Bundesbank (2021) and covers the period from 02.09.2019 until 01.11.2021.



(a) Bond prices

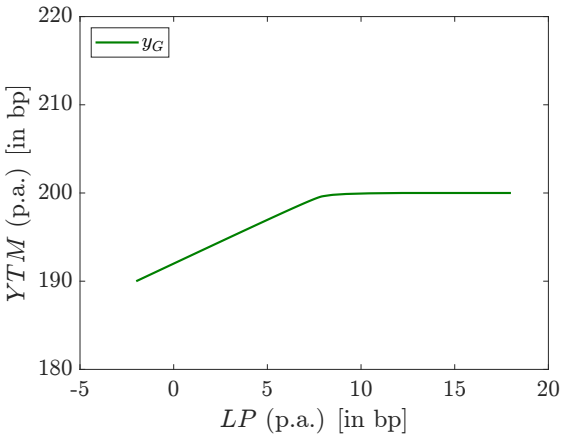
(b) Bond yields

Figure 7: Model results for different LP

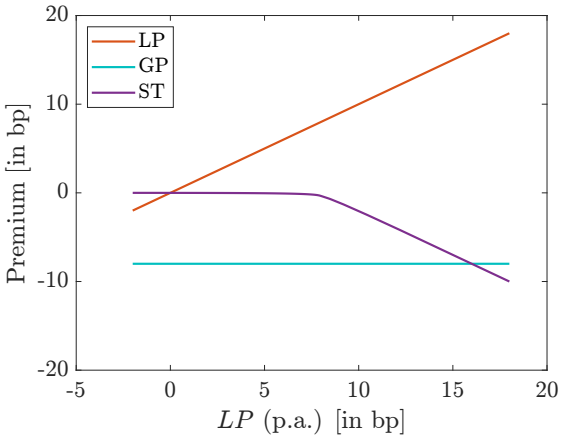
The model results displayed in the figures above are based on a green premium of $GP = 8$bp, a risk-free rate of $r_f = 200$bp, $\sigma = 0.0031$, $a = 11.919$, $T = 3.1$ years and a trinomial tree length of $N = 791$.

of the zero coupon bonds (i.e., FV=1), a lower price P_t, ceteris paribus, implies a higher yield to maturity y_t, and vice versa. The conventional bond C is assumed to be liquid, and thus not affected by changes in the premium that compensates for illiquidity of the asset. Therefore, the price P_C (see Figure 7a) and yield y_C (see Figure 7b) are unaffected by changes in the expected liquidity premium LP. On the other hand, the value of the illiquid green bond IG is affected as investors require a higher compensation for their liquidity risk and are thus only willing to pay lower prices. The German green bond G differs from the bond IG by having the additional switch option ST. This prevents the bond price P_G from assuming values lower than P_C. In the same fashion, the yields y_G cannot assume values higher than y_C. When the green premium outweighs the liquidity premium, the model yield y_G is smaller than the yield of the conventional bond y_C. This implies for the secondary market that a negative yield differential Δy (i.e., $\Delta y = y_G - y_C$) is observed. Finally, it should be noted that the green premium GP is equal to 8bp for all scenarios in the figure. For the German green bonds, this means that the yield difference Δy can be equal to zero, although there exists a green premium GP larger than zero. In such cases, liquidity effects dominate and the value of the upper threshold for the yield, y_C, is assumed. Moreover, in the case of bonds without a switch option, IG, the yield difference to a conventional twin can even assume positive values. This means that liquidity effects of bonds can potentially compensate the green premium. The model suggests that issuers and investors should therefore incorporate the bonds' exposure to illiquidity in their emission and valuation decision, respectively. This finding aligns with the published investor presentations or Green Bond Frameworks from France, Netherlands and Belgium, who all address liquidity aspects of their bonds (see section 3.2). Moreover, the model also shows that the German approach can prevent the yield spread from becoming positive. Therefore, it can be a viable method for issuers to mitigate by illiquidity induced risks.

Figure 8 displays the yield of a German green bond for different degrees of illiquidity and a decomposition of its yield premium that exists relative to its conventional twin. The green premium GP is constant with $GP = 8$bp for all values of the expected liquidity premium LP. The figure demonstrates that the switch option prevents the yield y_G to become larger than the yield of the conventional bond y_C. Further, its value (in bp) reflects the payoff structure of a short call



(a) Yield of a German green bond



(b) Decomposition of the yield

Figure 8: Model results for German green bonds

The model results displayed in the figures above are based on a green premium of $GP = 8$bp, a risk-free rate of $r_f = 200$bp, $\sigma = 0.0031$, $a = 11.919$, $T = 3.1$ years and a trinomial tree length of $N = 791$.

option on the illiquidity of the bond. Using this analogy, the strike would coincide with the value of GP. If LP assumes a value larger than GP, the switch option ST needs to compensate this difference.

The time series of the yield spreads displayed in Figure 3 indicate that a value of -5bp can be a realistic value for German green bonds. Based on the model results displayed in Figure 7b and Figure 8b, this would imply $LP = 3$bp and $ST = 0$bp, assuming $GP = 8$bp. In words, this model specification indicates that the greenium is sufficiently larger than the liquidity premium so that market intervention by the Finance Agency is very unlikely to be necessary and thus the value of the switch option is equal to zero.

5.2. Maximum Switch Option Value

Figure 9 shows that the yield of the German green bond y_G is capped by y_C at an expected liquidity premium LP that is larger than the green premium GP. This is because the stochastic liquidity premium might still assume a lower value, in which case the execution of the switch option, (i.e., the execution of switch transactions) is not optimal. Based on Equation 1, we know that at this point the difference $LP - GP$ coincides with the maximum value of the switch option $ST^{\max}$, as Δy is equal to zero. This maximum value is relevant as it indicates how much additional liquidity costs in excess of a greenium the holders of a German-type green bond can bear until they assign the same value to it as to a conventional government bond. In comparison, in the case of an illiquid green bond without the switch option, this value would be zero.

In light of the above, the maximum value of the switch option $ST^{\max}$ can be defined as,

$$ST^{\max} = \max_{LP} \{ST \mid y_G \leq y_C\}. \tag{24}$$

In the following, we provide an overview of how this measure changes for different model specifications and an estimation precision of 0.01bp. Table 5 shows the value of the switch option at execution, $ST^{\max}$ for different levels of GP. The model results show that $ST^{\max}$ is unaffected by the size of GP, ceteris paribus. This is because a higher GP increases the expected illiquidity LP that can be tolerated before the switch option is executed. From an issuers' perspective, this implies that by adopting the German approach, they can compensate an additional liquidity premium of 4.1bp compared to conventional green bonds until the yield differential Δy assumes a value equal to zero.

To put the value of 4.1bp into perspective, we assume a total issuance volume of 5 bn. EUR which equals the size of the smallest currently issued German green bond. This implies a potential maximum value of approximately 2 mn. EUR for the switch option, given an issuance volume of 5 bn. EUR. However, the Green bond from this example currently (01.11.2021) trades at a spread Δy of -8bp. Based on the model calibration displayed in Figure 7b, this would imply a $LP < 5bp$, for which the value of the switch option ST is equal to zero (see Figure 8b).

Table 6 shows the value of $ST^{\max}$ for different local volatilities of the underlying process for the liquidity premium. The model results indicate that a higher σ increases the maximum value of the switch option $ST^{\max}$. This is plausible, as a higher volatility of the stochastic liquidity premium increases the chance of realizing very low values, while larger values do not change the outcome once the threshold is y_C is reached. This means that the switch option is executed later, which implies a higher value for LP and $ST^{\max}$. The model also accommodates a special case assuming a non-stochastic liquidity premium. In this case, the option is executed for $LP = GP$. As the liquidity premium cannot change over time, the option is executed as soon as liquidity effects and the green premium cancel each other

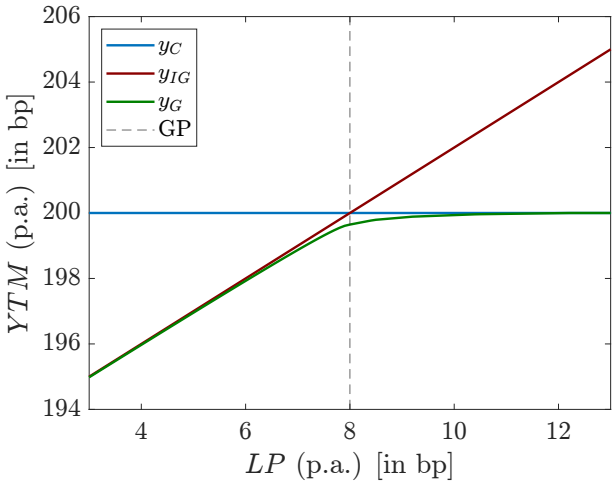


Figure 9: Bond yields for different LP

The model results displayed in the figures above are based on a green premium of $GP = 8$bp, a risk-free rate of $r_f = 200$bp, $\sigma = 0.0031$, $a = 11.919$, $T = 3.1$ years and a trinomial tree length of $N = 791$.

Table 5: Option value at execution for different GP

GP	LP	$ST^{\max}$
0	4.06	4.06
10	14.06	4.06
20	24.06	4.06
30	34.06	4.06

The table shows the values of $ST^{\max}$ for different GP based on a risk-free rate of $r_f = 200$bp, $\sigma = 0.0031$, $a = 11.9$, $T = 3.1$ years and a trinomial tree length of $N = 791$.

out. The resulting maximum value of the switch option before execution, $ST^{\max}$, is equal to zero in this scenario. For $LP < GP$, there is no chance that the option is executed as it implies a certain negative yield differential Δy. Therefore, the value of the option is equal to zero in this case as well.

Finally, Table 7 shows $ST^{\max}$ for different times to maturity T. In the model, this increases the length of the trinomial tree because $\Delta t = \frac{T}{N} = \frac{1}{250}$ is held constant. The results indicate a lower maximum value of the switch option $ST^{\max}$ for longer maturities T. This is explained by the decreasing likelihood of the stochastic liquidity premium realizing an outcome lower than LP. Therefore, the switch option is executed for a lower expected liquidity premium LP reducing its maximum value $ST^{\max}$.

5.3. Sensitivity Analysis of Bond Yields

The sensitivities of the initial yield to maturity to changes in the model parameters are estimated using finite differences that is motivated by a Taylor approximation. This approximation is required because a closed-form solution is not available due to the non-closed form of the model. Following Brandimarte (2006), a symmetric approximation of the first partial derivative of the yield y_0 with regard to the model parameters is computed, as this approach yields a lower order truncation error compared to forward or backward approximation. In its general form, the first derivative can be estimated using,

$$\frac{\partial y_0(x)}{\partial x} \approx \frac{y_0(x+h) - y_0(x-h)}{2h}, \tag{25}$$

where h denotes a small and constant value and x the parameter of interest, while the other model parameters are hold constant. The resulting sensitivities are displayed in Figure 10. The figures indicate that the sensitivity of the German green bond G has a continuous part, and a discontinuous part with jumps when LP assumes values above a certain threshold. The number of observed jumps in the figures for G coincide with $j_{max} = 4$ (or $-j_{min}$) of the calibrated model. One viable explanation might be that nodes in the tree switch to the value of the conventional bond, if the liquidity premium assumes a high enough value so that $P^G < P^C$ (see Equation 15). This also explains the continuous part on the left-hand side of the figures, as a switch scenario does not occur for low values of LP.

Figure 10a describes how much units the yield changes, if LP changes by one unit. The yield of the illiquid green bond y^{IG} changes by one basis point, if LP increases by one basis

Table 6: Option value at execution for different σ

σ	LP	$ST^{\max}$
0	8	0
0.002	10.53	2.53
0.004	13.28	5.28
0.008	18.81	10.81
0.010	21.26	13.26

The table shows the values of $ST^{\max}$ for different σ based on $GP = 8$bp, a risk-free rate of $r_f = 200$bp, $a = 11.9$, $T = 3.1$ years and a trinomial tree length of $N = 791$.

Table 7: Option value at execution for different T

T	LP	$ST^{\max}$
1	12.35	4.35
5	12.06	4.06
10	11.51	3.51
20	10.76	2.76
30	10.31	2.31

The table shows the values of $ST^{\max}$ for different T (constant Δt) based on $GP = 8$bp, a risk-free rate of $r_f = 200$bp, $\sigma = 0.0031$ and $a = 11.9$. Changes in T affect the tree length N, as Δt is hold constant with $\Delta t = \frac{T}{N} = \frac{1}{250}$.

point, while y^C is unaffected by changes in LP. The sensitivity of y^G ranges between 1 and 0. This aligns with the notion that the German green bond is valued as a conventional bond if LP is sufficiently high and valued as a counterfactual bond without switch option, if LP is sufficiently low, assuming a constant GP. In those cases, the stochastic process for LP either cannot assume values where y^G is lower than y^C, or where the switch option is executed. Figure 10b implies that a higher instantaneous volatility σ decreases y^G. This is because the downside potential is restricted by the switch option, while a lower realized liquidity premium reduces y^G. The parameter a describes the mean reversion rate of the stochastic process. Therefore, this sensitivity is inversely related with the sensitivity of y^G to σ. Finally, an increase in T, increases the yield y^G as well. Based on the absolute size of the sensitivities, the evaluation suggests that changes in σ and LP have the strongest impact on the model results. In light of the evaluation, it should be noted that the sensitivities only reflect the impact of small changes in the parameters. Further, their changes and thus the effect on the model results is restricted by their plausible range. Nevertheless, the model outcome might be significantly larger or smaller, if different estimates for those parameters are chosen.

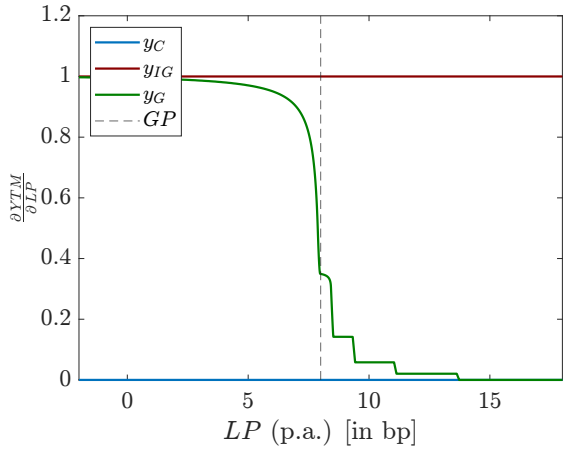
5.4. Limitations

The above discussed model for the green bond yields provides a first insight into the potential effects of the switch option between green and conventional bonds, which was pioneered by the German twin bond approach. However, the model is subject to some limitations that are discussed in the following.

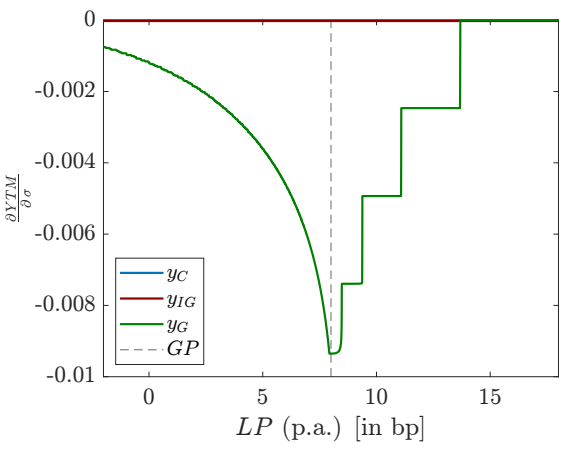
First, the model cannot decompose observed green bond yields $\hat{y}_G$ into the different components suggested by the model. Namely, the observed yield of the respective conventional twin $\hat{y}_C$, the liquidity premium LP, the green premium GP and the added-value of the switch option ST. This means that a calibration of the model parameters is not straightforward and proxies need to be applied instead. Moreover, this impedes the validation of the model results based on actual observations.

Another possible limitation can be the assumed process for the liquidity premium and its translation into a trinomial tree representation. For example, the Vasicek process in Equation 2 assumes a constant volatility and is, in addition to a mean-reversion parameter, defined by its first two moments. This means that it cannot accommodate possible volatility clusters or skewness that is introduced by jumps in the liquidity premium, as shown in Figure 6. Moreover, deriving the trinomial tree representation, we assume a maximum range from LP_{jmin} to $LP_{j_{max}}$ for the liquidity premium to ensure positive tree probabilities. This creates an upper and lower threshold that the liquidity premium cannot exceed. However, increasing the volatility of the process may provide a first idea of the possible implications when accounting for these effects, as it increases the overall dispersion of the stochastic premium.

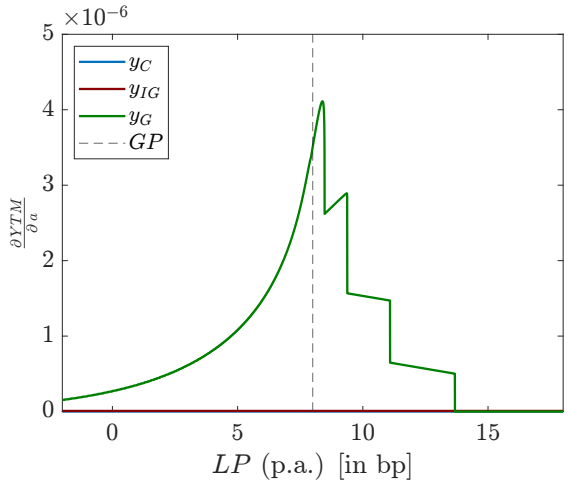
Finally, the model assumes a constant risk-free rate r and green premium GP. While adding additional complexity to the model by introducing more flexible (e.g., stochastic or



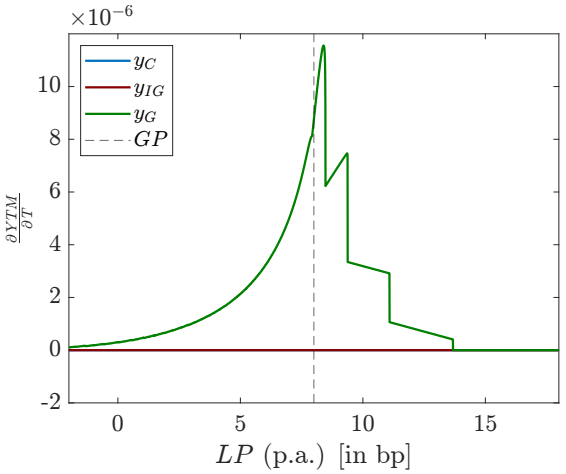
(a) Sensitivity to changes in LP



(b) Sensitivity to changes in σ



(c) Sensitivity to changes in a



(d) Sensitivity to changes in T

Figure 10: Model sensitivities

The model results displayed in the figures above are based on a green premium of $GP = 8$bp, a risk-free rate of $r_f = 200$bp, $\sigma = 0.0031$, $a = 11.9$, $T = 3.1$ years, a trinomial tree length of $N = 791$ and $h = 0.00001$

time-dependent) components might improve the calibration to observed yield spreads, this is not relevant for the main objective of this dissertation to better understand the potential impact of the switch option.

6. Conclusion

The goal of this dissertation is to provide a theoretical model for the pricing of green bonds that are based on the German twin bond approach. The focus here is on improving the understanding of the potential effects of introducing a switch mechanism between green bonds and their conventional counterparts. For this purpose, a non-closed form solution was derived that decomposes the yield differential into three effects: A liquidity premium, a green premium and the added value of the switch option. The model assumes a stochastic liquidity premium that follows a Vasicek process in discrete time, a constant green premium as well as a constant risk-free rate. The switch mechanism is modelled by assuming the theoretical value of conventional bonds as a lower limit for the green bond prices. For the model calibration the term structures of German Bundesanleihen and Pfandbriefen are used to obtain a proxy for the stochastic liquidity premium.

The main learning from the model is that the switch option can in certain conditions increase the value of the green bonds, which corresponds to a lower yield. Based on the calibration of the model, a maximum added-value of 4.1 bp before the execution of the option was identified. This translates to a maximum value of about 2 mn. EUR assuming a green bond with a 5 bn. EUR issuance volume. This means that issuers adopting the twin bond concept may be able to secure lower costs of capital compared to a traditional green bond concept that does not provide the switch option. For investors the concept reduces their exposure to potential liquidity risks by using the liquid conventional bonds to create a

lower limit for the green bond price. The model improves the understanding of the twin bond concept and thereby fills a gap in the literature. From a practical perspective, the model implications may assist issuers in the design choice of their green bond framework. For example, Denmark decided to adopt the twin bond concept, including a switch mechanism, which supports the potential benefits of this approach.

Green bonds are one important instrument to finance the transition to a more sustainable economy. In light of the significant growth of the green bond market in recent history and the competing frameworks, it is crucial to elaborate on their respective advantages and disadvantages. While this work contributes to the understanding of the twin bond switch mechanism, the current model can be further developed. On the one hand, an improved proxy for the liquidity premium and a larger sample of historic data may affect the calibration results, which can impact the size of the evaluated effects. On the other hand, a more sophisticated stochastic process for the liquidity premium and less restrictive assumptions in its discrete representation may increase the precision of the model results. In a broader context, one should evaluate if a high issuance volume of green bonds can affect the liquidity of similar conventional bonds, and whether a potential effect vanishes for lower volumes. If such effects are found, this would support the relevance of the twin bond approach with switch option to mitigate liquidity risks, as lower overall issuance volumes may be required. Otherwise, ensuring a critical volume that is high enough to avoid liquidity costs may be a viable alternative to this concept.

References

Almeida, M., Filkova, M., Harrison, C., & Sette, P. (2019). *Green bond european investor survey.* Retrieved from https://www.climatebonds.net/resources/reports/green-bond-european-investor-survey-2019 (Last accessed: 19.01.2022)

Bachelet, M. J., Becchetti, L., & Manfredonia, S. (2019). The green bonds premium puzzle: The role of issuer characteristics and third-party verification. *Sustainability*, *11*(4), 1098.

Bongaerts, D., & Schoenmaker, D. (2020). Green certificates: a better version of green bonds. *Policy Contribution 2020/20, Bruegel*.

Brandimarte, P. (2006). *Numerical methods in finance and economics*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Brigo, D., Dalessandro, A., Neugebauer, M., & Triki, F. (2009). A stochastic processes toolkit for risk management: Geometric brownian motion, jumps, garch and variance gamma models. *Journal of Risk Management in Financial Institutions*, *2*(4), 365–393.

Chen, L., Lesmond, D., & Wei, J. (2007). Corporate yield spreads and bond liquidity. *The Journal of Finance*, *62*(1), 119–149.

Climate Bonds Initiative. (2021). *Sustainable debt market summary q3 2021.* Retrieved from https://www.climatebonds.net/resources/reports/sustainable-debt-summary-q3-2021 (Last accessed: 19.01.2022)

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, *53*(2), 385.

Danmarks Nationalbank. (2022). *Green bonds.* Retrieved from https://www.nationalbanken.dk/en/governmentdebt/green_bonds/Pages/Default.aspx (Last accessed: 19.01.2022)

Deutsche Bundesbank. (2021). *Term structure of interest rates in the debt securities market - estimated values: Listed federal securities and pfandbriefe.* Retrieved from www.bundesbank.de (Last accessed: 19.01.2022)

Díaz, A., & Escribano, A. (2020). Measuring the multi-faceted dimension of liquidity in financial markets: A literature review. *Research in International Business and Finance*, *51*, 101079.

Dick-Nielsen, J., Feldhütter, P., & Lando, D. (2012). Corporate bond liquidity before and after the onset of the subprime crisis. *Journal of Financial Economics*, *103*(3), 471–492.

Dorfleitner, G., Utz, S., & Zhang, R. (2021). The pricing of green bonds: external reviews and the shades of green. *Review of Managerial Science*.

Doronzo, R., Siracusa, V., & Antonelli, S. (2021). Green bonds: The sovereign issuers' perspective. *Bank of Italy Markets, Infrastructures, Payment Systems Working Paper No. 3*.

Dutch State Treasury Agency. (2019). *Investor presentation green dsl.* Retrieved from https://english.dsta.nl/binaries/dsta-english/documents/publication/2019/05/06/investor-presentation-green-dsl/Investor+presentation+Green+DSL.pdf (Last accessed: 19.01.2022)

Ehlers, T., & Packer, F. (2017). Green bond finance and certification. *BIS Quarterly Review September*.

European Commission. (2012). *Charter of fundamental rights of the european union: 2012/c 326/02.* Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT (Last accessed: 19.01.2022)

European Commission. (2020). *Stepping up europe's 2030 climate ambition: Swd/2020/176 final.* Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020SC0176 (Last accessed: 19.01.2022)

European Commission. (2021). *Proposal for a regulation of the european parliament and of the council on european green bonds: Com/2021/391 final.* Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0391 (Last accessed: 19.01.2022)

Fama, E. F., & French, K. R. (2007). Disagreement, tastes, and asset prices. *Journal of Financial Economics*, *83*(3), 667–689.

Finanzagentur GmbH. (2020). *Green bond framework.* Retrieved from https://www.deutsche-finanzagentur.de/fileadmin/user_upload/institutionelle-investoren/pdf/GreenBondFramework.pdf (Last accessed: 19.01.2022)

Finanzagentur GmbH. (2021a). *The bund's green twins: Green federal securities.* Retrieved from https://www.deutsche-finanzagentur.de/en/institutional-investors/federal-securities/green-federal-securities/ (Last accessed: 19.01.2022)

Finanzagentur GmbH. (2021b). *Green bond investor presentation september 2021.* Retrieved from https://www.deutsche-finanzagentur.de/fileadmin/user_upload/institutionelle-investoren/pdf/Green_Bond_Investor_Presentation_2021_II.pdf (Last accessed: 19.01.2022)

Finanzagentur GmbH. (2021c). *Joint press release 20 april 2021: Federal government publishes first allocation report for green federal securities.* Retrieved from https://www.deutsche-finanzagentur.de/fileadmin/user_upload/pressemeldungen/en/2021/2021-04-20_pm03_allocation_report_2020_en.pdf (Last accessed: 19.01.2022)

Finanzagentur GmbH. (2021d). *Primary market.* Retrieved from https://www.deutsche-finanzagentur.de/en/institutional-investors/primary-market/ (Last accessed: 19.01.2022)

Fisher, L. (1959). Determinants of risk premiums on corporate bonds. *Journal of political economy*, *67*(3), 217–237.

Fuller, W. A. (2009). *Introduction to statistical time series*. John Wiley & Sons.

Gianfrate, G., & Peri, M. (2019). The green advantage: Exploring the convenience of issuing green bonds. *Journal of Cleaner Production*, *219*, 127–135.

Goldstein, M. A., Hotchkiss, E. S., & Pedersen, D. J. (2019). Secondary market liquidity and primary market pricing of corporate bonds. *Journal of Risk and Financial Management*, *12*(2), 86.

Hachenberg, B., & Schiereck, D. (2018). Are green bonds priced differently from conventional bonds? *Journal of Asset Management*, *19*(6), 371–383.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton Univ. Press.

Harrison, C., & Muething, L. (2021). *Sovereign green, social, and sustainability bond survey.* Retrieved from https://www.climatebonds.net/files/reports/cbi-sovereign-green-social-sustainability-bond-survey-jan2021.pdf (Last accessed: 19.01.2022)

Hartzmark, S., & Sussman, A. (2019). Do investors value sustainability? a natural experiment examining ranking and fund flows. *The Journal of Finance*, *74*(6), 2789–2837.

Hayashi, F. (2000). *Econometrics*. Princeton, NJ: Princeton Univ. Press.

Helmcke, S., Heuss, R., Hieronimus, S., & Engel, H. (2021). *Net-zero deutschland: Chancen und herausforderungen auf dem weg zur klimaneutralität bis 2045.* Retrieved from https://www.mckinsey.de/de/news/presse/studie-net-zero-deutschland-klimaneutralitaet-chancen-herausforderungen (Last accessed: 19.01.2022)

Helwege, J., Huang, J.-Z., & Wang, Y. (2014). Liquidity effects in corporate bond spreads. *Journal of Banking & Finance*, *45*, 105–116.

Hull, J. (2018). *Options, futures, and other derivatives* (Ninth edition, global edition ed.). Harlow, England: Pearson.

Immel, M., Hachenberg, B., Kiesel, F., & Schiereck, D. (2021). Green bonds: shades of green and brown. *Journal of Asset Management*, *22*(2), 96–109.

International Capital Market Association. (2021). *The green bond principles: Voluntary process guidelines for issuing green bonds.* Retrieved from https://www.icmagroup.org/sustainable-finance/the-principles-guidelines-and-handbooks/green-bond-principles-gbp/ (Last accessed: 19.01.2022)

ISS ESG. (2020). *Second party opinion (spo): Sustainability quality of the issuer and the green german federal securities.* Retrieved from https://www.deutsche-finanzagentur.de/fileadmin/user_upload/institutionelle-investoren/pdf/SPO.pdf (Last accessed: 19.01.2022)

Kapraun, J., Latino, C., Scheins, C., & Schlag, C. (2019). (in)-credibly green: Which bonds trade at a green bond premium? *Proceedings of Paris December 2019 Finance Meeting EUROFIDAI - ESSEC*.

Kempf, A., Korn, O., & Uhrig-Homburg, M. (2012). The term structure of illiquidity premia. *Journal of Banking & Finance*, *36*(5), 1381–1391.

Kempf, A., & Uhrig-Homburg, M. (2000). Liquidity and its impact on bond prices. *Schmalenbach Business Review*, *52*(1), 26–44.

Koziol, C., & Sauerbier, P. (2007). Valuation of bond illiquidity. *The Journal of Fixed Income*, *16*(4), 81–107.

Löffler, K. U., Petreski, A., & Stephan, A. (2021). Drivers of green bond issuance and new evidence on the "greenium". *Eurasian Economic Review*, *11*(1), 1–24.

MacAskill, S., Roca, E., Liu, B., Stewart, R. A., & Sahin, O. (2021). Is there a green premium in the green bond market? systematic literature review revealing premium determinants. *Journal of Cleaner Production*, *280*, 124491.

Monfort, A., & Renne, J.-P. (2014). Decomposing euro-area sovereign spreads: Credit and liquidity risks*. *Review of Finance*, *18*(6), 2103–2151.

People's Bank of China. (2021). *Green bond endorsed projects catalogue (2021 edition)*. Retrieved from https://climatebonds.net/market/country/china/green-bond-endorsed-project-catalogue (Last accessed: 19.01.2022)

Rendleman, R. J., & Bartter, B. J. (1980). The pricing of options on debt securities. *The Journal of Financial and Quantitative Analysis*, *15*(1), 11.

République Française. (2017). *Framework for the green oat*. Retrieved from https://www.aft.gouv.fr/files/archives/attachments/25562.pdf (Last accessed: 19.01.2022)

République Française. (2021). *Investors presentation - march 2021*. Retrieved from https://www.aft.gouv.fr/files/medias-aft/3_Dette/3.2_OATMLT/3.2.2_OATVerte/Investor%20presentation%20AFT%20GIC%20March%202021_UK.pdf (Last accessed: 19.01.2022)

Riedl, A., & Smeets, P. (2017). Why do investors hold socially responsible mutual funds? *The Journal of Finance*, *72*(6), 2505–2550.

Schestag, R., Schuster, P., & Uhrig-Homburg, M. (2016). Measuring liquidity in bond markets. *Review of Financial Studies*, *29*(5), 1170–1219.

Schuster, P., & Uhrig-Homburg, M. (2012). The term structure of bond market liquidity conditional on the economic environment: An analysis of government guaranteed bonds. *No 45, Working Paper Series in Economics, Karlsruhe Institute of Technology (KIT), Department of Economics and Management*.

Schwarz, K. (2019). Mind the gap: Disentangling credit and liquidity in risk spreads. *Review of Finance*, *23*(3), 557–597.

SIFMA. (2021). *2021 capital markets fact book*. Retrieved from https://www.sifma.org/resources/research/fact-book/ (Last accessed: 19.01.2022)

Svensson, L. (1994). Estimating and interpreting forward interest rates: Sweden 1992-4. *C.E.P.R. Discussion Papers*, *1051*.

The Kingdom of Belgium. (2018). *Green olo framework*. Retrieved from https://www.debtagency.be/sites/default/files/content/download/files/green_olo_framework.pdf (Last accessed: 19.01.2022)

United Nations. (n.d.). *United nations sustainable development goals*. Retrieved from https://sdgs.un.org/goals (Last accessed: 19.01.2022)

United Nations. (2022). *Chapter xxvii environment 7. d paris agreement*. Retrieved from https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27&clang=_en (Last accessed: 19.01.2022)

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, *5*(2), 177–188.

Wegener, C., Basse, T., Sibbertsen, P., & Nguyen, D. K. (2019). Liquidity risk and the covered bond market in times of crisis: empirical evidence from germany. *Annals of Operations Research*, *282*(1-2), 407–426.

Wulandari, F., Schäfer, D., Stephan, A., & Sun, C. (2018). The impact of liquidity risk on the yield spread of green bonds. *Finance Research Letters*, *27*, 53–59.

Zerbib, O. D. (2019). The effect of pro-environmental preferences on bond prices: Evidence from green bonds. *Journal of Banking & Finance*, *98*, 39–60.

Junior Management Science 7(3) (2022) 690-730

Junior Management Science

journal homepage: www.jums.academy

Opportunities and Challenges in Commissioning Materiality-Driven Sustainability Reporting Towards the SDGs: The Case of Cadeler A/S

Darian Nestor Weicht

University of Southern Denmark

Abstract

Frequently and recently tightening and expanding sustainability reporting policies and requirements can pose significant administrative burdens on SMEs upholding a strong culture of accountability to their stakeholder network. This seminal case study examines how a Danish offshore wind farm commissioner can efficiently (1) navigate towards credibility in and (2) derive actionable insights from their sustainability (reporting) integration trajectory by capitalizing on the increasingly emphasized materiality principle. Group-based Fuzzy AHP and Textual Analysis aim to excavate and assess senior managers' and external stakeholders' preferences based on the GRI Standards and the UN's SDG targets. Internal priorities emphasize safety, compliance, and profitability, whereas external stakeholders' and their groups' priorities exhibit mixed findings on their type and extent of alignment with the former. Content elements assigned higher relative importance tend to be more robust to changes in decision-makers' uncertainty and verbal bias. The author confirms that a simplicity-informativeness trade-off tends to be driven by stakeholder grouping and that a data-driven, subject-based, and objectifying approach should be complemented with context, managerial judgment, and process iteration.

Keywords: Sustainability; materiality; prioritization; credibility; actionability.

1. Introduction

Burgeoning streams within natural sciences stress increasingly adverse anthropogenic influence reflected in growing negative impacts on ecosystems' planetary boundaries through depletion of groundwater, rare earth metals and fossil fuels, and increasing GHG release (Bebbington et al., 2019). In the late 1990s, the Dutch chemist and Nobel laureate Paul Crutzen coined the term "Anthropocene" to describe the geological epoch in which human behavior is the primary driver of increasingly drastic changes in environmental processes, behaviors, and livelihoods (ibid.; Crutzen, 2006; Zoomers & Otsuki, 2017). In a parallel movement to worsening ecological conditions, conceptually similar trends in the social realm of human society gave birth to CSR, which was devised to hold organizations accountable for increasing social tension in communities and encourage a transition towards corporate citizenship that is argued to have emerged with the inception of the Industrial Revolution (Carroll, 2009; Schwartz & Carroll, 2003; Wren, 2005).

During the 1960s, social responsibility shifted from a remote phenomenon into a corporate matter and expanded to community affairs (ibid.). Scholars started defining CSR, with one side emphasizing that businesses should use their resources in the interest of society and an intimate relationship between the two (e.g. Frederick, 1960), and another questioning the legitimacy of behavior that contradicts economic rationale (Friedman, 1970). Following conceptual acceleration in the 1970s, research streams from 1980s call for voluntary CSR adoption through businesses to maximize its effectiveness and focus on their stakeholder audience (Freeman, 1984; Jones, 1980). First empirical studies on stakeholder theory have evolved in the 1990s with CSR activity experiencing increasing adoption in Europe and the US (Carroll, 2009; Moon, 2005), which will be successful only if (1) mainstream organizations engage in CSR and incentivize related activity and (2) such commitment has a tangible and positive impact on the organization by adding to their environmental and/or social bottom line (Vogel, 2006).

Synchronically, such environmental catastrophes as the Exxon Valdez oil spill in Alaska in 1989 that arose from failing governance mechanisms and lacking sensitivity to

DOI: https://doi.org/10.5282/jums/v7i3pp690-730

ESG and reputational risks (Bebbington, Larrinaga, & Moneva, 2008; Maguire, 2010) have raised the need for organizations to credibly explain to an increasingly pressuring stakeholder audience how their business model and activity creates value over time and how it tackles sustainability challenges through innovation in technology and/or management control; this type of transparency has become a critical requirement for their long-term survival (Bebbington & Larrinaga, 2014; Bocken, Short, Rana, & Evans, 2014; Busco, Frigo, Quattrone, & Riccaboni, 2014; Dyllick & Muff, 2016; International Integrated Reporting Council, 2013). The incident gave birth to the CERES Principles that allow investors and stakeholders to assess an organization's environmental performance (Orsato, 2009). Since the 1980s, the link between the occurrence of manmade natural disasters and the emergence of so-called Green Clubs that entail a wide range of guidelines, charters and programs to help organizations manage their reputational risk at cost of voluntary entry has intensified (ibid.; Bebbington et al., 2008; Power, 2009).

As part of the trend towards stakeholder centricity, legitimacy, and reputational capital, SR emerged as a complement to the corporate reporting portfolio, which was primarily focused on an organization's financial health and, unlike its nonfinancial counterpart, has gained substantial credibility, consistency, comparability, and maturity over the course of the 20th century that was shaped by conflicts, market crashes, and accounting scandals (Abernathy, Stefaniak, Wilkins, & Olson, 2017; Baron, 2014; Rupley, Brown, & Marshall, 2017; Tschopp & Huefner, 2015). Along with a substantial increase in investors' and public attention to nonfinancial performance, reporting on such has gained significant uptake and has become a standard routine for multinational firms in response to increasing stakeholder demand (Amir & Serafeim, 2018; KPMG, 2020).

To an extent, this development can also be attributed to expansion in international agreements and frameworks such as the UN's SDGs and the UNGC that provide companies with extensive guidance on improving SR (Jespersen & Olmsted, 2019; Rasche, Gwozdz, Larsen, & Moon, 2020; United Nations, 2015a, 2015b). This carries particular importance for increasingly mandated SR that emerged from South Africa as part of the King Codes stressing more stakeholder-centric corporate governance and caried over to the NFRD that mandates SR for larger-scale EU-based organizations in accordance with the UNGC's principles that encourages thorough sustainability risk management, performance management and measurement, and stakeholder engagement (Baron, 2014; European Commission, 2014; KPMG, 2016; Rasche et al., 2020).

With earlier studies doubting the value of SR (Burritt & Schaltegger, 2010; Vinnari & Laine, 2013), an increasing number of studies points at a wide range of (non-)financial advantages adopters can have over non-adopters and can therefore capitalize on dual returns arising from linking business-as-usual activity with strategic CSR (Baron, 2014; Dyllick & Muff, 2016). To start with, the value-generating ability of the organization can be enhanced by reaping such direct benefits as better-informed decisions related to strategic direction, improved (opportunity) cost management, reduced coordination effort by building common ground or gains in productivity or reputational capital (Jespersen & Olmsted, 2019). Transparent reporting specifically can mitigate risk factors' impact on the organization and improve their access to cheaper debt and equity financing (Abernathy et al., 2017; Cheng, Ioannou, & Serafeim, 2014). Indirect benefits include, but are not limited to, higher legitimacy, greater chances of taking supplier roles as a result of higher transparency and information-processing capacity, and thus competitive edges from superior business relationships, project contracts, and improved access to capital markets (Jespersen & Olmsted, 2019).

With the advantages tending to outweigh a number of potential shorter-term drawbacks that can arise from greater transparency, it seems as if the adoption of nonfinancial reporting seems straightforward, especially since a growing number of investors and nonfinancial stakeholders tend to attribute financial value relevance to such disclosure practice (Barman, 2015; Grewal, Hauptmann, & Serafeim, 2020; Schiehll & Kolahgar, 2020). With a broad variety in approaches and high-level regulation with virtually absent enforcement mechanisms (Johansen, 2016), a major challenge resides in the implementation of SR that provides concise coverage on topics that matter to statement users without rendering the preparation of such reports into a comprehensive compliance exercise (Power, 2009). Put differently, the present lack of binding comprehensive regulation and guidance that would discourage cherry-picking and predominantly aspirational talk tends to disincentivize potentially committed organizations to identify, prioritize, and report on sustainability topics that are genuinely material to the business and their stakeholder's informational needs (Beske, Haustein, & Lorson, 2020; Christensen, Morsing, & Thyssen, 2013).

Very recently, the NFRD's review and the introduction of the EU Taxonomy on Sustainable Finance, along with continuously growing market interest, tend to push reporting organizations from greenwashing and towards more data-driven and comprehensive SR to efficiently direct capital flows to sustainable investment projects that are likely to yield dual returns for impact and ESG investors (Blowfield & Murray, 2008; Dyck, Lins, Roth, & Wagner, 2019; EU Technical Expert Group on Sustainable Finance, 2019; European Commission, 2014). With the EU Taxonomy becoming a mandatory disclosure that institutional investors are mandated to adhere to by late 2021, organizations outside financial services are required to follow suit by late 2022 (ibid.). With a majority of stakeholders involved in the NFRD review requesting that (1) the policy apply to organizations that are privately owned and count at least 250 members and (2) a tighter integration of sustainability performance indicators and higher-quality and more comprehensive explanations on which topics matter most and how they relate to the organization's sustainability strategy, an increasing reporting burden would start

applying to SMEs that may be constrained in committing organizational resources towards SR and therefore need to channel their capacity towards an approach that allows them to engage in informative disclosure practice, most notably through the identification of key nonfinancial topics and recipients (Baumann-Pauly, Wickert, Spence, & Scherer, 2013; Germanwatch, 2021; Majoch, Hoepner, & Hebb, 2017).

To illustrate how a smaller-sized organization can effectively move towards SR that is informative about their key sustainability drivers, this thesis report develops a possible point of departure for Cadeler A/S, a Copenhagen-based offshore windfarm commissioner that very recently became a public-interest entity on the OSL and very likely to become subject to tightening SR regulation in the wake of the organization's high-pace growth trajectory. With the EU heavily promoting renewable energy and committing up to 800 billion euros to offshore windfarm development plan as part of the Green Deal sealed in 2019, Cadeler A/S is assumed to substantially benefit from high-quality SR and strong financials to build trust with ESG investors and increase the likelihood to attract more financial resources to accelerate the Green Transition towards a carbon-neutral EU economy over the next decade (Cheng et al., 2014; EU Technical Expert Group on Sustainable Finance, 2019; European Commission, 2020a, 2020b; La Torre, Sabelfeld, Blomkvist, & Dumay, 2020; United Nations, 2015b).

This study focuses on how the organization can credibly substantiate their mission statement of contributing towards a sustainable future by means of a structured materiality-driven approach that fosters transparent SR beyond compliance and towards competitive edges (Churet & Eccles, 2014; Orsato, 2006). Apart from releasing their first SD Report in April 2021 as part of the SR mandate imposed through the firm's listing, the organization has a largely absent track record with SR and is expected to face significant competitive disadvantages, should they refrain from improving disclosure quality (ibid.). In particular, reporting against the SDGs has become a commonplace reporting activity and tends to be linked with thorough implementation of a recognized SR framework (Global Reporting Initiative, 2020a; Pizzi, Rosati, & Venturelli, 2021). Cadeler's unique situation with the opportunity to make a substantial contribution towards greening the EU's energy mix as part of the Green Deal's agenda motivates the following research question:

> "What are the opportunities and challenges that Cadeler A/S should be wary of when integrating materiality-driven ESG Reporting that is more stakeholder-centric and in line with de facto established SR Standards the latest EU legislation on nonfinancial disclosure?"

The author attempts to answer this question geared towards effective SR and related information processing by answering two sub-questions in this report:

1. How could Cadeler A/S prioritize material sustainability topics when planning the production of an SD Report that signals compliance with established frameworks and regulation, alignment with industry practices, and contribution towards relevant SDGs?
2. What are the steps to take for Cadeler A/S when integrating materiality-driven SR into organizational routines, management practices, and respective and sustainability control systems?

The first question deals with the concept of materiality and its dual nature that can drive both transparency and cherry-picking when preparing a value relevance-driven SD report. With the Standards released by the GRI in 2016 as the framework's sixth iteration, expanded by additional reporting standards in early 2021, and representing a de facto established SR framework, the GRI Standards represent a powerful framework that helps a reporting organization to live up to legal SR requirements imposed on the regional level due its strong link to the UN's SDGs, the UNGC, the NFRD, and ties to IR coming from an attempt to harmonize frameworks that presently work complementarily (Buhmann, 2018; Villiers & Maroun, 2017; La Torre, Sabelfeld, Blomkvist, Tarquinio, & Dumay, 2018; Sonnerfeldt & Pontoppidan, 2020).

The second question draws on the latest development preparing compliant SR with such frameworks as the GRI Standards, the NFRD, or the EU Taxonomy that is bound to substantiate SR practice by imposing sector-specific sustainability criteria for economic activity is intended to point at a possible action plan that bring Cadeler A/S closer to more substantiated future SR. This plan includes increasing information-processing efficiency and potential synergies from jointly using management control and sustainability control systems (George, Siti-Nabiha, Jalaludin, & Abdalla, 2016). Despite some studies proposing a transition from internal controls to external reporting, the opposite pathway can also produce high-quality SR that does not necessarily prioritize a feel-good narrative over stating genuine sustainability performance development, risks, and opportunities (Alrazi, Villiers, & Staden, 2015; Derchi, Zoni, & Dossi, 2020; Figge, Hahn, Schaltegger, & Wagner, 2002; Traxler, Schrack, & Greiling, 2020). Throughout the report, the author takes an information-processing perspective on how a resource-constrained organization like Cadeler A/S can capitalize on the materiality concept and leverage its potential towards truthful (self-)reporting to keep key recipients well-informed and increase financial markets' efficiency (Orlitzky, 2013).

Like in every scientific study, delimitations apply to this report. To start with, the paper is primarily concerned with prioritizing material topics with a quantitative case study approach and their implications on the construction of an external nonfinancial reporting system and how this can potentially impact its adoption an integration process. The study assumes that the applied SR frameworks are structurally efficient and do not induce perverse incentives, though those shortcomings may apply to specific sectors or industries. Likewise, the study is rather suggestive on the topics reported or those that could be disclosed outside of applied

frameworks and is reserving the assessment of framework-induced performance measure congruity to emerging research branches. The thesis is, to the author's knowledge, one of few reports that empirically test a structured materiality assessment methodology outside of the study it was applied to at inception (e.g. Calabrese, Costa, Levialdi, & Menichini, 2016) and is intended to provide the case study company, Cadeler A/S, with hands-on implications on how to prioritize SR topics in a controlled setting and pinpoint opportunities and challenges in its planning, adoption, and construction to effectively integrate materiality-driven SR on the organizational level. Therefore, the thesis departs from existing policies and suggests a possible baseline towards developing unique screening criteria, metrics or distinctive reporting topics for the renewables sector which are deemed advisable once an informative SR foundation is in place.

The remainder of the thesis report is organized as follows. Section 2 reviews literature on accounting and reporting for SD and introduces the concept of materiality as a key concept in this report, and outlines benefits and challenges inherent to adopting such routines rendering related processes and disclosures informative and credible for investors and non-financial stakeholders. Section 3 describes the methodological approach taken to analyze and draw implications from sustainability topics that are potentially material to Cadeler A/S and assumed to be informative on the organization's sustainability-related efforts and effects. Section 4 presents the results obtained from the internal and external materiality assessment and their corresponding SDGs. Section 5 discusses the results and their broader implications for Cadeler A/S and concludes with limitations and suggestions for future research.

2. Theoretical Background

2.1. Nonfinancial Reporting

2.1.1. Developments in the Reporting Landscape

Prior to the rise of stakeholder theory and SR, communication to the stakeholder audience was primarily directed to providers of financial capital through financial statements and primarily looked at reporting organizations through lenses of profitability, productivity, and risk management (Tschopp & Huefner, 2015). Its history traces back to the early 1900s with its value to the user increasing during the Progressive movement in the 1920s and 1930s, increasing industrialization, and more frequent and extensive access to capital markets (ibid.). With local developments and refinements in GAAP, the transition towards IFRS as a global reporting framework pushed forward by such intergovernmental institutions and NGOs as the OECD to enable global comparability of organizational financial performance and the IFRS' harmonization with potentially stricter local regulations poses ongoing challenges, as does the dichotomy of rules-based and principles-based regulation (ibid.).

Over time, the corporate reporting portfolio expanded substantially from brief income statements to complementary financial statements, risk management and governance disclosures, and standalone SRs to give report users a clearer picture of an organization's past, present, and future (non-)financial performance and risks and their drivers, partly in response to the formation of CERES in response to the Exxon Valdez oil spill in 1989 (Baron, 2014; Rupley et al., 2017). Along with the advent of extensive financial reporting regulation and increasing stakeholder pressure in partial response to failures to hold organizations accountable for (non-)financial misconduct yet primarily for improved decision-making (Rupley et al., 2017; Tschopp & Huefner, 2015), both financial and nonfinancial disclosure have become increasingly comprehensive and sparked a discussion on how to render corporate disclosure more concise without sacrificing informativeness to the comprehensiveness-conciseness tradeoff outlined in the (non-)financial reporting literature (Jespersen & Olmsted, 2019). To illustrate, Stolowy and Paugam (2018) underline SR's significant uptake which tends to grab relatively more investor attention than financial information, implying that providers of financial capital tend to put nonfinancial reporting under more scrutiny to improve their asset allocation towards dual returns (Agrawal & Hockerts, 2019).

In 2013, the IIRC presented IR as a potential solution that integrates financial and nonfinancial strategic considerations into six capitals that the organization's business model uses as input factors for long-term value creation which tend to represent the outputs it generates. The underlying idea is that a higher input factor quality, such as more stringent focus on hard factors such as financial performance management paired with internal policies improving soft factors such as human and relational capital, can improve the states of either of the capitals and use the realized benefits as refined input factors to further improve their output (International Integrated Reporting Council, 2013). The (shared) value added to outputs by focusing on the inputs factor quality is assumed to accumulate over time an is contingent on the content elements that have the most material influence on a firm's business model's ability to create value (ibid.; Porter & Kramer, 2011).

Taking together the most prominent examples on SR, at least five organizations including the GRI, UNGC, PRI, IIRC, EC, and ISO provide reporting guidance, with most reporting frameworks serving as complements with partial overlap and therefore (1) representing a substantial administrative burden to organizations aspiring to truthful reporting and (2) creating the need to streamline nonfinancial reporting guidance and standards to reduce complexity and render SR more accessible, comparable, consistent, and reliable (La Torre et al., 2018). However, most frameworks encourage voluntary adoption and the NFRD, which tends to be well-received as an initial move towards mandating nonfinancial reporting on the EU level, does not require the reported information to be independently reviewed and can encourage organizations to ramp up disclosure volume by employing a comply-or-explain clause reducing the density of decision-useful information and likely incurring excess opportunity cost (European Commission, 2014; Johansen, 2016; La Torre et al.,

2020). Insights obtained from its review towards more credible SR implying a certain degree of unsatisfaction with the stringency of the NFRD's present state reveal that stakeholder pressure and the call for stronger enforcement will likely intensify in the near future and make a substantial contribution to the administrative reform building common ground for more evidence-based policymaking and more comparable, reliable, and consistent SR to support the efficient allocation of capital flows towards genuinely sustainable investments to bring SR on par with financial accounting and reporting (EU Technical Expert Group on Sustainable Finance, 2019; Germanwatch, 2021; R. Gray, 2006; La Torre et al., 2020; Lucarelli, Mazzoli, Rancan, & Severini, 2020).

2.1.2. The GRI Standards

The concept of sustainability or rather SD presently lacks a superior definition and can cause confusion when combined with growth aspirations (Robinson, 2004) and implications from SR that are contingent on the framework chosen for such (Wu, Shao, & Chen, 2018). The GRI originated in 1997 as a collaboration project of CERES and the Tellus Group in Boston and released the first version of their SR framework in 2000 and has gone through five major iterations between 1997 and 2016 (Jebe, 2019; Matuszyk & Rymkiewicz, 2018) though lacks a definition of SD either yet guides the document user towards a "three-dimensional model based on an organization's economic, environmental and social impact" (Villiers & Maroun, 2017:4). The framework geared towards holistic and informative ESG reporting aims to establish a common language between companies and stakeholders through the provision of reporting principles, ruling on general (voluntary) disclosures, and guidance on how to deal with sustainability issues spanning across economic, environmental, and social topics (Villiers & Maroun, 2017).

Despite the framework's aspiration to guide a reporting company towards more informative SR by stressing the importance of disclosing material content elements that have a significant impact on the business and its stakeholders, its semantics tend to make the G4 appear as a framework that is encouraged to be adopted by organizations with sufficient available financial, temporal, and cognitive resources (Calabrese et al., 2016). With further reviews stating that these guidelines tend to suffer from partial redundancy, lack of clarity, and causing perceived overwhelm to reporting organizations, the GRI Standards, the framework's latest iteration and in effect since 1st July 2018, render SR more accessible in several ways, which are outlined below.

First, changing the semantic title from "Guidelines" to "Standards" signals a higher degree of legitimacy and tends to set them more on par with established financial accounting standards. Second, this transition is further restructuring the framework into modules (Global Reporting Initiative & University of Stellenbosch Business School, 2020). That is, the GRI Standards 101, 102 and 103 set the framework's foundations and general baseline SR requirements. Economic, environmental, and social topics are split over 34 topic-specific standards and 89 disclosures indicating how sub-elements of a topic shall be measured and reported (ibid.). Coverage on each topic is voluntary and encourages reporting based on what is deemed most material to the organization and their stakeholders; omissions, however, must be justified through a comply or explain clause (Pizzi, Venturelli, & Caputo, 2020). Reporting organizations can then decide whether to report on all management approach disclosures and at least one topic-specific disclosure of a material topic (core) or all of them (comprehensive) (ibid.). Its modular structure also enables more responsive SR updates and encourages quick adaptation to changes in compliance requirements (Skouloudis, Evangelinos, & Kourmousis, 2009) and provides a "balanced disclosure on management, economic, environmental and social sustainability themes" (Wu et al., 2018:1).

Second, the GRI Standards refine and emphasize the concept of materiality to the catalogue and therefore organizations making or intending to make optimal use of the framework by a thorough materiality analysis outlined in GRI 102-46 and GRI 102-47 (Machado, Dias, & Fonseca, 2021; Global Reporting Initiative & University of Stellenbosch Business School, 2020). Third, its rule-based nature, structured construction and detailed guidance on compliant reporting and metrics makes it a convenient alternative for organizations who are used to rule-based routines inherent to financial reporting and want to refrain from principles-based regulation (Reynolds & Yuthas, 2008). Finally, its strong stakeholder focus encourages truthful, simple, and appropriate reporting through its aspiration to balance comprehensiveness and conciseness when adhered to and makes it responsive to changes in stakeholders and their informational requirements (ibid.). Presently, the GRI framework tends to be the preferred alternative among organizations planning to adopt SR (Rodríguez-Gutiérrez, Guerrero-Baena, Luque-Vílchez, & Castilla-Polo, 2021) and their database lists 63,852 Reports and 38,484 GRI Reports (as of 22nd May 2021).

2.1.3. IR and SR: Mutually Exclusive Concepts?

SR is a critical first step for organizations to assess the extent to which their activities contribute to planetary (un-)sustainability and identify where to be held accountable for sustainability impacts, potentially through cooperation with investors and the public sector (R. Gray, 2006). Put differently, an organization can increase its chances of long-term survival by transparently engaging with their stakeholders and conveying the link between their business model activities and value creation over time (International Integrated Reporting Council, 2013). Value, however, does not share a common definition and perception in individuals and thus needs to be reified to resolve disagreement on its presence and type of contribution (Bourguignon, 2005). For instance, shared value, coined by Porter and Kramer (2011) and created by organizations through profit-making and stakeholder engagement geared to social welfare, is a relatively novel concept and is thus earlier in its reification process and pre-

sumably more susceptible to be challenged for conceptual shortcomings (Bourguignon, 2005). The failure to recognize compliance or challenges to reconcile an organization's financial profit-orientation with that directed towards nonfinancial surpluses and thus breaking a perceived trade-off tends to be a general issue brought up in academic literature since the inception of shareholder theory, stakeholder theory, and CSR (Crane, Palazzo, Spence, & Matten, 2014; Friedman, 1970).

Intertwined with value creation and relevant sustainability topics is their linkage to the organization's business model, which consists of three pillars essential to gaining and sustaining competitive advantage (Barney, 1991): (1) value proposition, i.e. the incremental benefit customers are only receiving at the providing organization and are willing to devote monetary resources to, (2) value delivery, i.e. how organizational stakeholders receive incremental benefits, and (3) value capture, i.e. how the providing organization retains the incremental benefit received (Richardson, 2008). Bocken et al. (2014) extend this traditional perspective by introducing eight generic business model archetypes geared towards the creating of sustainable value through a technological, social, and/or organizational focus. With technological business models focusing on material and/or energy efficiency, circularity and green substitutes, socially oriented archetypes emphasize functionality instead of ownership, stewardship, and frugal behavior, whereas business models with an organizational focus tend to be geared towards so centering around societal needs and scalability (ibid.). Since both the GRI and IR framework require disclosure on the properties of the organization's business model, focusing on such can render SR more credible when aligned with highly material content elements (Global Reporting Initiative & University of Stellenbosch Business School, 2020; International Integrated Reporting Council, 2013; Morioka, Evans, & Carvalho, 2016).

Whereas IR explicitly stresses value creation over time, the GRI is more explicit on sustainability impacts induced by a reporting organization (Global Reporting Initiative, 2020a; International Integrated Reporting Council, 2013). Despite a growing stream of research building on the value-impact distinction and further conceptual difference that are assumed to lead to the adoption of either alternative (e.g. Jensen & Berg, 2012), recent findings signal a degree of harmonization by identifying increasing overlaps between the GRI's SR framework and IR (Petcharat & Zaman, 2019). Therefore, one can argue that GRI-based reporting is not exclusively impact-oriented albeit its emphasis on value creation over time is less explicit (ibid.). Implications from this finding are twofold: First, distinguishing between SR and IR may become less clear in the future given complementarities between reporting styles (ibid.) and when factoring in efforts directed towards the harmonization of the wealth of nonfinancial reporting frameworks in addition to IR and GRI, such as SASB's guidance and a number of assurance standards such as the AA1000AS (Jebe, 2019; La Torre et al., 2018; Saenz, 2019; Safari & Areeb, 2020). Second, IR with its partial integration into the GRI Standards becomes more accessible to organizations that face higher reporting capacity constraints and more informative given informational complementarities between SR and IR (ibid.), since the former in the form of GRI tends to be well-adopted in reporting organizations worldwide (e.g. Rupley et al., 2017).

2.1.4. Disclosure Quality and Enforcement

A burgeoning stream of research notes that SR quality that can be influenced by such internal organizational drivers and external forces as mandated SR policies embedded in government regulation (Badia, Bracci, & Tallaki, 2020; Mies & Neergaard, 2020). With respect to non-mandated IR, recent investigations assert that such reports tend to be inflated and susceptible to increases in disclosure volume and impression management when organizational earnings are low (Melloni, Caglio, & Perego, 2016). Similarly, Stacchezzini, Melloni, and Lai (2016) reveal biased disclosures and a positive link between sustainability performance and the extent of information provided on such along with an overall substantial lack of quantification.

Du and Yu (2020) find that future CSR performance tends to be significantly affected by the disclosure material's readability and textual sentiment. It is likely that the attempt to conceal sub-par sustainability performance at the expense of disclosure quality will divert investor interest from publicly listed organizations engaging in such practice (H. Z. Khan, Bose, Mollik, & Harun, 2020). Therefore, it can be inferred that emphasis on underlying impacts and effects can positively impact perceived credibility from (non-)financial stakeholders (ibid.). Factoring in concerns on SR frameworks' integrity, it seems as if disclosure-related opportunism applied to nonfinancial reporting frameworks tends to occur independently of such. As a potential remedy, Amran, Lee, and Devi (2014) recommend to tightly integrate CSR management into organizational processes and governance mechanisms to legitimize SR and disclosure credibility on the firm's vision and mission statement, potentially along with external stakeholder pressure.

Regarding GRI report quality in the wind energy sector, Moseñe, Burritt, Sanagustín, Moneva, and Tingey-Holyoak (2013) show that institutions do not necessarily enforce a transition towards higher-quality SR sufficiently. In the case of the Spanish Wind Energy sector, Iberdrola as a "strategic company leader" (ibid.:210) has capitalized on their unique ability to prepare extensive SR documentation in accordance with the GRI framework and their size to put "coercive pressure" (ibid.) on their industry rivals that went beyond that imposed by institutions. As a result, Iberdrola paved the way for how Spanish organizations operating in the Wind Energy sector disclose sustainability information based on GRI by imitating the organization (ibid.). Adopting nonfinancial reporting and the degree of framework compliance conveyed by organizations operating in the wind energy sector is a largely unexplored field, making this thesis report one of the first of its kind to assess which SR topic tend to be most prioritized and how pronounced GRI compliance in reports is

reflected.

To the author's best knowledge, only one longitudinal study by Talbot and Boiral (2018) assesses the degree to which organizations in the energy sector adhere to the requirements laid out in the GRI framework. The authors find that 92 percent of audited sample reports and the practices from six out of 21 organizations tend to engage in non-compliant SR and tend to apply a diverse methodology of impression management techniques (ibid.). Disclosure of non-compliant information spans from lacking transparency in the calculation methodology applied to GHG emissions over incomplete information on environmental metrics to content that lacks entirely (ibid.). Over time, the authors find mixed yet mainly negative evidence on improvements in environmental SR and changes in content elements of SR disclosure material and attribute the results to two mechanisms (ibid.). First, implemented legitimation strategies tend to downplay the lack of impact data that may have significant implications on the organization's sustainability strategy and emphasize commitment to future disclosure of presently lacking information (ibid.).

Second, there tends to be a positive link between non-compliant reporting and an organization's propensity to distort balanced disclosure by artificially inflating positive news and further downplaying their negative counterpart through deliberate neglect of information or transferring negative information to less visible spots in reports such as footnotes (ibid.). Likewise, organizations tend to manipulate numbers, for instance by replacing total GHG emissions with GHG intensity or changing reference points when reporting environmental performance data (ibid.). Importantly, the conclusions hold for organizational SRs that were approved by an independent third party that presumably enhances the reports' credibility to the stakeholder audience (Porter & Kramer, 2011). Therefore, one can argue that stakeholders with a less sophisticated assessment tools are likely to have difficulty in verifying SR information (Talbot & Boiral, 2018). Notably, the study relies on the GRI G3, which tend to be less refined than the GRI Standards in that they do not reflect 10 years of learning since the integration of the materiality principle in the GRI framework and evolution in assurance (Abernathy et al., 2017; Brown, Jong, Lessidrenska, & Mossavar-Rahmani, 2007; Gerwanski, Kordsachia, & Velte, 2019; Jebe, 2019; Matuszyk & Rymkiewicz, 2018).

Given the rather pessimistic review of institutional influence on SR, what has changed in the last years and where will it go? Over the last decade and most notably eight years, the call for putting a tighter grip on organizations and investors has been translated into the Accounting Directive 2013/34/EU and, more importantly, the NFRD which can be considered the first mandatory nonfinancial reporting regulation introduced on the EU level (Johansen, 2016). It has recently been complemented by the EU Taxonomy framework to fuel the transition towards a carbon-neutral economy and achievement of the UN's SDGs' ambitions laid out in the 2019 EU Green Deal (EU Technical Expert Group on Sustainable Finance, 2019). The NFRD presently requires organizations with more than 500 employees to report on four pillars based on the UNGC's principles, namely (1) social & employee-related matters, (2) environmental matters, (3) respect to human rights, and (4) anti-corruption and bribery (European Commission, 2014). According to § 19a, each of these pillars requires a description of the organizations business model, related policies, their outcome and risks to manage, and which metrics are used to capture performance on the respective pillar (ibid.).

Although an integrated comply-or-explain clause and a rather open policy design are intended to foster informative SR reporting and organizational learning towards it (Buhmann, 2018), the NFRD does not require an assurance engagement to verify the information disclosed (ibid.). Along with opening the door to managerial opportunism in CSR disclosure due to the ambiguity inherent to nonfinancial information, the optionality of external verification implies that its adoption can be seen as beyond-compliance behavior though making it a statutory requirement can have a positive impact on SR quality and enhance its credibility (Junior, Best, & Cotter, 2014; Porter & Kramer, 2011). Mandating such verification procedures and more rule-based reporting tend to positively impact SR quality and performance, which tend to be substantially driven by cultural factors and the sophistication and experience of a country's legal system in setting SR rules (Ioannou & Serafeim, 2012; Mies & Neergaard, 2020; Scholtens & Sievänen, 2013) and temporal factors (Conway, 2019). Denmark, France and Norway, for instance, have introduced mandatory CSR reporting in 2009 and 2001 and 2013, and well ahead of other EU members though do not prescribe any specific SR framework (Baron, 2014). Although mandatory reporting has a significantly positive impact on preserving environmental and social conditions (Eccles, Krusz, & Serafeim, 2012), such standardization can render SR compliance-driven, make adopters neglect potential benefits, and could "portray business-as-usual as genuine efforts in ESG" (Baron, 2014:26). The introduction of the EU Taxonomy on Sustainable Finance together with an extensive review on the NFRD, tend to underline the EU's transition towards more standardized and mandated SR activity (Baron, 2014; EU Technical Expert Group on Sustainable Finance, 2019; Germanwatch, 2021; Matteo La Torre et al., 2020), with the former elaborated on in the following.

The EU Taxonomy on Sustainable Finance represents the centrepiece of the EU's efforts to channel capital flows towards activities that sustainably develop the economy towards the UN's SDGs and carbon-neutrality by 2030 and 2050, respectively, by setting sector-specific technical screening criteria (TSC) that classify an economic activity as sustainable and thus able to contribute to environmental policy objectives (EU Technical Expert Group on Sustainable Finance, 2019; European Commission, 2020b). The TSCs are tailored to a wealth of economic activities defined by the NACE Rev 2 framework intend to ensure that the activities channel efforts towards six environmental objectives, namely (1) climate change mitigation, (2) climate change adaptation, (3) sustainable use of water and marine re-

sources, (4) transition to a circular economy, was prevention and recycling, (5) pollution prevention and control, and (6) protection of healthy ecosystems (ibid.). According to the framework, an economic activity is environmentally sustainable (i.e. Taxonomy-aligned) when it (1) makes a substantial contribution towards at least one of the six objectives and (2) does no significant harm to any of the six targets, (3) complies with minimum social safeguards outlined by the ILO core labor conventions, and (4) adheres to the activity-specific technical screening criteria (ibid.). There are presently TSCs for climate change mitigation and adaptation objectives and on how an adopting organization shall avoid harm to the remainder, indicating that parts of the framework are under very recent development (ibid.).

Identifying reporting requirements can complement environmental reporting based on the GRI Framework since such catalogues as the EU Taxonomy, which is virtually the first evidence-based SR policy, tend to stimulate a significant negative relation between such policies made and GHG emissions on a supranational level (Lucarelli et al., 2020). This framework will require Cadeler A/S to substantiate their environmental reporting by the extent to which their realized Revenue and incurred Capex and Opex materialize through environmentally sustainable activity and contribute to decarbonizing the EU economy. Furthermore, it is assumed to enhance the organization's ability to attract capital flows from ESG investors and convey its theory of change on how capital flows build an economically and environmentally desirable outcome proxied by the SDGs (EU Technical Expert Group on Sustainable Finance, 2019; United Nations, 2015a, 2015b). The following section emphasizes the materiality principle, which will, along with other desirable characteristics such as timeliness and comparability, be assigned special emphasis in this thesis report. The underlying idea is that (1) the goal of SR should be the provision of decision-relevant information towards the organization's stakeholder audience, and (2) incorporating the materiality principle into organizational processes can reduce long-term reporting and communication costs to the organization.

2.2. Materiality Analysis: Engaging Stakeholders with a Double-Edged Sword

Present the wealth of benefits that an organization can reap with the integration of sustainability accounting, reporting and increased sensitivity towards accountability, how can it start its transitional journey towards comprehensive sustainability integration (Dyllick & Muff, 2016) and implement a nonfinancial reporting ecosystem that is in line with recognized SR frameworks and the latest (upcoming) regulatory practice in the EU? In the words of Torelli, Balluchi, and Furlotti (2020), "Materiality is the driver through which companies can select issues to be included in nonfinancial reports favouring the expectations of all stakeholders" and the key ingredient for integrating ESG matters into organizational routines and refining a firm's sustainability strategy towards the creation of shared value (Whitehead, 2017). In other words, incorporating the materiality principle into organizational processes can reduce long-term SR and communication costs and foster innovation and beyond-compliance behavior (ibid.). Nonetheless, SR informativeness is inherently multi-dimensional and is presumed to hinge on how well an adopting organization can credibly identify and prioritize the focus areas of its sustainability strategy and external SR (e.g. Torelli et al., 2020).

Considered "one of the cornerstones of accountancy" (Frishkoff, 1970:116), the materiality principle plays a dual role in that (1) it sets the minimum extent of disclosure reporting organizations their auditors need to verify and (2) it sets boundaries on disclosure-grade elements to ensure concise and balanced reporting that does not expend on potentially competitive edges (Baumüller & Schaffhauser-Linzatti, 2018; Hsu, Lee, & Chao, 2013). Consequently, it channels an organization's financial, temporal, and cognitive resources towards informative disclosure practice by maximizing their efficiency and effectiveness absent both (1) overstraining those resources' capacity and (2) misdirected reporting efforts towards virtually immaterial content elements (G. L. Gray, Turner, Coram, & Mock, 2011). Importantly, unnecessary strain on statement users' information-processing capacity is avoided and reduces information asymmetry in the organization-stakeholder dyad (Churet & Eccles, 2014; Jebe, 2019). On the EU level, the materiality principle is embedded in the Accounting Directive 2013/34/EU and the NFRD which guide the preparation of financial and nonfinancial statements (Baumüller & Schaffhauser-Linzatti, 2018). Similarly, Jebe (2019) notes that laying relevance thresholds over a reporting firm tends to ensure their compliance to externally imposed policies and therefore (1) a license to operate, and (2) an increased likelihood to attract interest from sophisticated impact investors providing capital flows (Deegan, 2002; Weber, 2016).

2.2.1. Organizational Drivers

Recent empirical investigations note that a technical-rational approach applied to common materiality assessments is not free of managerial judgment and tends to objectify its inherent subjectivity and should therefore be complemented by socio-political considerations and dialogic accounting which is virtually synonymous with active stakeholder engagement (Puroila & Mäkelä, 2019; Zhou, 2011). Specifically, the former note that materiality assessment tends to entail considerate variability across (1) underlying objectives, (2) authorities of information, (3) potential (in-)comparability, and (4) transferability (ibid.). Depending on either a society- or corporate-centered approach materiality assessments can either be used to make sustainability considerations or stakeholder relationship management part of the organization's decision-making agenda or expose risk factors to the business and its finances or stakeholders' informational needs (ibid.; Puroila, Kujala, & Mäkelä, 2016). Authorities of information tend to be defined on a continuum ranging from a strictly internal corporate to a strictly external and group-based matter reflected in methodologies

falling in the same continuum (ibid.). Regarding criteria used to determine material topics, technic-rational criteria are represented by materiality matrices or co-measurement across categories, whereas topic-specificity favors evaluating topics to defined and isolated categories such as economic, environmental, or social; importantly, results obtained from different approaches may suffer from incomparability (ibid.).

Finally, the assessment's outcome can differ in its nature and therefore tends to limit its transferability to changes in the reporting organization's operational or strategic financial and/or sustainability context (ibid.; Gerwanski et al., 2019). Put differently, a truthful and accurate snapshot of an organization's material sustainability issues may lose its ability to capitalize on these two traits over time and tends to be susceptible not only to changes in the reporting organization's operational and strategic focus, but also to the composition of its stakeholder audience that tends to use such assessment as a guidance to assess the organization's strategy towards dual returns (ibid.; Beske et al., 2020). Conversely, it is likely that inaccurately disclosed material elements become accurate over time though unlikely for an external statement reader to identify with limited organizational knowledge (Talbot & Boiral, 2018).

2.2.2. Benefits and Challenges

Typically, material topics are conveniently communicated through a matrix that captures their relative impact on the reporting firm and importance to their stakeholders (Bellantuono, Pontrandolfo, & Scozzi, 2016). In a burgeoning stream of research on this method's sensibility, Puroila and Mäkelä (2018) find that materiality disclosure of large-cap organizations with GRI-compliant SR tends to assume that reportedly material topics apply to the organization's entire stakeholder audience and is unclear on their relevance to each party. Put differently, a materiality matrix tends to over-aggregate and -simplify potentially complex stakeholder perspectives and assume comparability and (com-)measurability in their type and magnitude, meaning that disclosed overviews tend to be silent on stakeholders' and the organization's underlying economics and preferences (ibid.) and present "a compromise of different [divergent] perceptions on what sustainability information is material" (Puroila & Mäkelä, 2019:1056). Omitting detailed information on how the disclosed materiality matrix was constructed can question its credibility and necessitates the materiality assessment to be a structured and transparent procedure (Calabrese et al., 2016; Machado et al., 2021). Importantly, interpreting omissions through a materiality lens are deemed imperative to not misinterpret information gaps in SR material (Unerman & Zappettini, 2014).

Besides improved stakeholder engagement and exposing business risks and value drivers, a materiality analysis tailored to the organization's core capabilities can be used to capitalize on organizational learning on such key outputs as GHG emissions, energy management, or safety measures to get a better grasp of how an organization's underlying processes drive their sustainability performance and use the insights as strategic inputs for the organizations' sustainability strategy and communicating risk factors to the stakeholder audience (Jebe, 2019; Puroila et al., 2016). In turn, resulting outputs serve as inputs for a refined analysis that can further build organizational legitimacy and work against the trend of instrumentalizing materiality analysis for impression management rather than credible SR and reframe the concept (Beske et al., 2020; Puroila et al., 2016). From an investor perspective, a transparent and concise disclosure on material nonfinancial items tends to positively impact an organization's stock price informativeness, value relevance and market- and firm-level ESG performance, where the market gains be explained by growing investor interest towards materiality-driven SR and accountability (Grewal et al., 2020; M. Khan, Serafeim, & Yoon, 2016; Schiehll & Kolahgar, 2020). Further benefits include improved investment risk management and lower opportunity cost through stronger focus on a market's underlying value drivers (Jebe, 2019). Plus, an organization and their stakeholders tend to face lower communication and agency cost in communicating their aspiration levels through a market, framework, or institutional lens (Cerbone & Maroun, 2020).

When adopting materiality-driven reporting, an organization tends to face several challenges and dilemmas. To start with, the concept of materiality, similar to sustainability, does not share a universal and/or mandated definition and what is perceived as material tends to be the result of social engineering driven by society's or addressees' context and framing (Jebe, 2019). Consequentially, different parties can describe the same issue in different terms with different perceived meanings shaped through an interplay of construction and intervention to such, which ultimately leads to a frame or rather perception of an element and makes an assessment inherently subjective (ibid.; Reimsbach, Schiemann, Hahn, & Schmiedchen, 2020). Further complexity is introduced by the absence of standardized methodology and guidance in nonfinancial reporting frameworks (Machado et al., 2021; Puroila & Mäkelä, 2019). Materiality-driven SR can be based on two types of materiality, namely (1) quantitative materiality (investor-oriented) of figures that are strongly informative about an organization's underlying economics, and (2) qualitative materiality (stakeholder-focused), which emphasizes information that is not economic or financial in nature yet can have financial implications on the organization's value-generating ability (Grewal et al., 2020; Jebe, 2019).

From another SR perspective, Zhou (2011) points at the dilemmas of conciseness and comprehensiveness and selective and mechanized reporting. Comprehensiveness is argued to result from increasing informational needs from stakeholders that are required to be reflected in previously concise reporting material, though other stakeholders may consider the same information less relevant to their decision-making (ibid.; Jebe, 2019). With diverse information demand from stakeholders, reporting organizations tend to transition from selective reporting focusing on key issues to mechanized reporting, which tends to be SR that is more streamlined, comprehensive, and compliance-driven

yet tends to lack topic prioritization and responsiveness to emerging topics that may be interrelated or challenge the value and belief system reported (ibid.). Compliance, on the other hand, tends to be relevant only if misreporting is sufficiently disincentivized to render materiality-driven reporting more cost-efficient and potentially enforceable (ibid.; Beske et al., 2020). A possible step is to interlink quantitative and qualitative materiality more tightly by (1) disincentivizing impression management and (2) separating financial and nonfinancial policymakers to co-evolve both systems to a comparable standard (Beske et al., 2020; Jebe, 2019).

2.2.3. Materiality in the GRI Framework

The GRI Standards take a multi-stakeholder perspective on SR and thus stress the principle of including external stakeholders in assessing material content elements to report that transcends through a four-step procedure (Global Reporting Initiative, 2013; Global Reporting Initiative & University of Stellenbosch Business School, 2020). The materiality principle states that SR must "reflect the reporting organization's significant economic, environmental, and social impacts; or substantively influence the assessments and decisions of stakeholders" (Global Reporting Initiative & University of Stellenbosch Business School, 2020:10) by setting relevance thresholds to ensure compliance (Jebe, 2019). In the assessment, the first three steps prior to SR disclosure entail (1) identifying, (2) prioritizing, and (3) validating material topics to the reporting organization and their stakeholder audience that qualitatively give an idea on how to allocate resources committed to SR; the final step comprises the review of the report under considerations of stakeholder inclusiveness and the organization's sustainability context for the next period's SR (Global Reporting Initiative, 2013).

The GRI Disclosures 102-46 and 102-47 require an organization to describe the approach through which they determined material sustainability topics and a list of such (Global Reporting Initiative & University of Stellenbosch Business School, 2020). Although moving beyond compliance by adhering to voluntary guidelines tends to be associated with positive feedback from regulators, assurance providers, and the broader stakeholder audience (e.g. Pizzi et al., 2021; Porter & Kramer, 2011), doing so tends to be an imperfect remedy against excess reputational or sustainability risk and also stakeholders with higher-powered informational needs. A practical example is represented by how both the organization's most salient stakeholder audience and most material topics are determined (Beske et al., 2020; Mitchell, Agle, & Wood, 1997; Torelli et al., 2020).

In very recent studies devoted to assessing disclosure quality of GRI-compliant SR, Machado et al. (2021) investigate how transparently a sample of GRI-compliant organizations report on determining key addressees and material topics as their SR's foundation. They find that most organizations tend to provide a high-level outline rather than a detailed description on their methods' assumptions and procedures (ibid.). This has three implications. First, a statement user interested in the methodology could question the credibility of the report with imperfect knowledge on the methodology provided by such. Second, a number of studies, statement users, and ESG rating agencies taking SR information as inputs are at risk of taking materiality and thus credibility for granted by assessing GRI compliance and adoption alone (e.g. Torelli et al., 2020). These findings hint at a positive association between lacking enforcement on transparent methodological disclosure and cherry-picking and incomparability (Puroila & Mäkelä, 2019).

Third, assurance providers and regulators carry not only responsibility in ensuring accuracy in more elaborate reporting demanded as part of the NFRD review, but also in that the organization has run a structured approach for determining SR content that was selected due to its relative importance rather than its convenient reportability (Beske et al., 2020; Calabrese, Costa, Ghiron, & Menichini, 2019; Germanwatch, 2021). Importantly, future iterations of the GRI framework could not only focus on broader coverage of transferable content elements, but also on providing more methodological guidance on determining material topics and salient stakeholder groups to improve comparability and accuracy of SR information, build congruence in the perception of materiality, incorporate the principle in a larger number of mandating policies such as the NFRD and the EU Taxonomy, and co-evolve policies with financial reporting and disincentivize cherry-picking in SR (Baumüller & Schaffhauser-Linzatti, 2018; Beske et al., 2020; Jebe, 2019; Machado et al., 2021; Puroila & Mäkelä, 2019).

2.3. Challenges in Sustainability Adoption and Integration

Integrating sustainability into daily decision-making tends to stimulate organizational learning and change processes that should be thoroughly coordinated, tailored to the organization's aspiration and promote proactivity (Buhmann, 2018). In a similar vein, Alrazi et al. (2015) find that firm-level sustainability performance tends to be driven by interrelation of an organization's perceived legitimacy, accountability, and proactivity. Specifically, accountability tends to be determined by the perceived informativeness of their SR and/or performance, whereas proactivity, the firm's ability and/or willingness to "invest in environmental management and accounting systems and stakeholder engagement [and affect their satisfaction]" (ibid.:44), drives accountability and SR's legitimacy. To enable successful sustainability integration in that CSR is used credibly (e.g. Koep, 2017), two overarching ideas should be considered. First, transitioning towards more sustainable operations is ideally phased over time to capitalize on learning economies and a growing knowledge repository to improve future decision-making, especially in the wake of an ambitious growth trajectory (Dyllick & Muff, 2016; Vermeulen & Barkema, 2002). Second, replicating a historically successful integration process that is not company-specific and will likely not yield the expected complementary benefits (Witjes, Vermeulen, & Cramer, 2017). Therefore, organizations are encouraged to identify a best practice that fits their unique portfolio of processes, routines, and aspirations to maximize the benefits

from their custom kind and degree of sustainability integration (Maas, Schaltegger, & Crutzen, 2016).

2.3.1. Raising Awareness Towards the SDGs

Unlike scientific coverage on the evolution and possible trajectories of nonfinancial reporting, empirical studies on the adoption of ESG reporting are emerging and tend to speak in favor of adopting SR. For instance, Rosati & Faria (2019) posit that early adoption of SDG reporting is more likely if a reporting organization has younger DMs and is strongly committed towards informative disclosure practice. Similarly, Pizzi et al. (2021) and Venturelli et al. (2020) assert that adopting the GRI Standards for SDG-oriented reporting tends to have a positive impact on performance on achieving SDG targets and that both the GRI and the SDGs tend to be intimately linked. Nonetheless, reporting performance tends to be influenced by how well a reporting organization adopts a prioritization of content elements and metrics that link the targets and underlying economics (Machado et al., 2021). Organizations should beware that credible reporting on sustainability performance is an inherently multidimensional reporting challenge and requires a holistic approach to the exercise (Adams & Larrinaga, 2019; Buniamin, Nazli, & Ahmad, 2015).

Further requirements allude to the presence of sustainability in business as the mere adoption of a SR framework tends to be insufficient and can raise credibility concerns (Jong, Harkink, & Barth, 2018). In a similar vein, Hallstedt, Ny, Robèrt, and Broman (2010) list three key generic requirements to increase the chances of successful ESG integration. First, nonfinancial aspirations must be tightly integrated into the organization's business-as-usual goals and plans. Second, those decision-making bodies with significant impact on the aspirations' achievement need to be provided (dis-)incentives to foster goal-oriented decision-making and a more efficient allocation of personal cost (Feltham & Xie, 1994). To illustrate, Derchi et al. (2020) suggest that CSR performance be tightly linked to manager compensation, provided that the MCS in place credibly captures the former. The author assumes that a realization of such may be achieved by introducing GRI metrics into the organization's performance evaluation system to support decision-making towards higher disclosure transparency and a stronger accountability to the stakeholder audience.

Finally, the authors speak in favor of adopting decision-support tools to both improve information-processing and decision-making and rendering incentive management more cost-efficient to make a tangible impact towards the advancement of societies and advancing business practices by creating a feedback loop between observant researchers providing evidence-based guidance to practitioners, who in turn provide inputs to academia (Bebbington & Larrinaga, 2014; Hallstedt et al., 2010). For instance, an organization committing to reporting on environmental impact tends to investigate their environmental cost in terms of (1) how they arise in a manufacturing/service provision process and (2) how environmental cost are captured and accounted for, often distinguishing by private cost to the organization from cost to the social context the organization is embedded in (Deegan, 2005). Private costs are usually represented by that of inputs resulting in by-product disposed or those costs incurred through the excessive and inefficient use of input factors (ibid.). These properties make the quantification and measurement of these private cost relatively straightforward, whereas public cost have a more indirect and complex nature that makes them difficult to capture (ibid.). Steering incentives towards a more refined cost-tracing mechanism may potentially unlock eco-efficiency gains and competitive edges that add sustainable value (Orsato, 2009; Schaltegger, Burritt, & Petersen, 2017). This, in turn, can be achieved by the skillful deployment of metrics and targets that channel SD efforts towards the SDGs (Buonocore et al., 2019).

2.3.2. Mobilizing (Top) Management

A rich body of literature focusing on the importance of management in adopting ESG-oriented routines and integrating them into existing workflows. For instance, Kiesnere and Baumgartner (2019) posit that support from top management and a responsive organizational culture are two key imperatives to successful sustainability integration. The former in particular must be willing and show commitment to allow linkages between organizational routines, culture, and SD concerns (ibid.; Adams & Frost, 2008).

Importantly, SD integration tends to require a "power promoter" (Kiesnere & Baumgartner, 2019:1607) that drives the integration process from the top or the bottom layers in the organizational hierarchy, assuming there is willingness and responsiveness to organizational change. Taking a top-down perspective, Adams and McNicholas (2007) identify several knowledge-driven impediments that managers across different levels of seniority can encounter: lacking knowledge on best practice SR and/or linking SD with the organization's strategy, unclarity on how report on SD and related KPIs, and indecisiveness on which reporting framework to use as guidance and a potential incapacity to separate financial, economic, or other nonfinancial KPIs. Metrics in particular have gained significant importance in organizational strategic planning, decision-making, and reporting yet are susceptible to comparability issues due to variability in input and output data used for calculation (Adams & Frost, 2008; Talbot & Boiral, 2018). Thus, shortcomings in managerial competency are likely to impede bottom-up integration mechanisms (ibid.).

Eccles, Ioannou, and Serafeim (2014) note that effective sustainability integration is realized by a high degree of formalizing CSR-oriented routines and by increasing top management's incentive sensitivity to it by making a larger fraction of their pay contingent on sustainability metrics; holding top management more accountable for nonfinancial performance tends to enable more extensive stakeholder management, a propensity to longer-term strategic commitments, and a higher degree of SR disclosure, of which especially environmentally related information tends to be influenced positively by the share of MBAs with a legal background (Ma,

Zhang, Yin, & Wang, 2019) yet less pronounced than social disclosures, similar to economic disclosures (Alshehhi, Nobanee, & Khare, 2018). In addition, a strong market position and commitment from its stakeholder community can increase an organization's accountability and legitimacy within their industry and financial market environment (Rodrigues & Franco, 2019).

A financial markets lens takes a confirmative role by observing a positive relation between sustainability commitments and significant long-term outperformance in the stock market pointing at a "virtuous circle" between the prioritization of CSR matters and an organization's economic prosperity, and investors' commitment to contribute (Eccles et al., 2014; Oikonomou, Yin, & Zhao, 2020:14; Winkler, Etter, & Castelló, 2020). Furthermore, studies confirm that public listing and consequential exposure to an expectant and diverse stakeholder community tends to increase internal pressure to engage in CSR reporting; the most notable impact tends to be visible on the top management level due to their alleged ability to drive sustainability integration most effectively (Thoradeniya, Lee, Tan, & Ferreira, 2015). The integration process, however, tends to be unintuitive and can be either split between economic, environmental and social impacts or tackled in an integrated way; it seems likely that either type of sustainability issues has to be assessed with a unique approach and unique timing indicative of prioritization (Morrison-Saunders & Therivel, 2006).

2.3.3. Evolving Internal Controls

Somewhat linked to management-related challenges are those pertaining to internal controls, which are intended to direct managerial decision-making towards the most efficient possible outcome (Feltham & Xie, 1994). An adopting organization should have an internal control system in place that can accurately capture financial and nonfinancial performance in the form of dedicated metrics to assess incentive alignment in the principal-agent relationship or rather ties between the organization's stakeholder audience and their management (Davidson, 2011; Figge et al., 2002; Whitehead, 2017).

Effectively integrating sustainability metrics into a control system that is intimately linked to manager compensation tends to have substantial influence on rendering organizational decision-making, internal processes, and governance codes more stakeholder-centric (Derchi et al., 2020; Rubino & Napoli, 2020). Kerr, Rouse, and de Villiers (2015) note that integrating SR into an MCS greatly supports the communication on and of such metrics and how beliefs and aspirations towards sustainability should be formalized, suggesting a BSC as a well-established and simple-to-use medium. Importantly, the authors find that a focus on (moving beyond) environmental compliance is positively linked to boundary systems such as certifications to assure lawful business conduct and stronger integration of environmental performance metrics and issues into their diagnostic and interactive controls, respectively, preferably indicators used for SR activity (ibid.). Assuming truthful intentions, organizations engaging in SR could benefit by integrating GRI metrics into their sustainability-driven performance measurement system and improve their data collection and progress monitoring mechanisms (Hubbard, 2009; Traxler et al., 2020) and also their progress towards SDG achievement (Buonocore et al., 2019).

Managerial motivation that can be based on incentive provision tends to have substantial influence on the design of both MSCs and SCSs and ultimately the SR system integrated in the organization (Herremans & Nazari, 2016). Notably, neither of them tends to be well-developed when SR is primarily compliance-driven (ibid.). However, strategic use of all three mechanisms can be achieved by learning across vertical and horizontal boundaries to differentiate the organization and their SR beyond compliance and industry rivalry when both control systems and SR practice are formalized relative to universally accepted SR guidelines. This implies that although navigating a diverse stakeholder community and industry peers and a company's own requirements and aspirations in search of unique best practices is resource-intensive and requires mechanisms to restrain managerial opportunism and foster informed and committed decision-making towards sustainability, it allows a thoroughly integrating, controlling, and reporting organization to advance their competitive positioning in the pursuit of a beyond-compliance strategy and transition from mechanistic to organic reporting that can potentially shape reporting practice on the industry level (ibid.; Moseñe et al., 2013).

Three additional remarks should be made. First , an organization does not necessarily have to agree with their stakeholder community in scope and scale of responsibility towards all CSR topics to become an industry leader in SR (Herremans & Nazari, 2016). Second, a reporting organization can substantially benefit by (1) building linkages between disclosed SR and its impact on their members' work across hierarchy levels and organizational performance and (2) assigning the compilation of the SR to representatives of several departments to obtain a holistic perspective on the past, present, and future context in which the organization has performed and aspires to create sustainable value (ibid.; Rodríguez-Olalla & Avilés-Palacios, 2017). Finally, board members deemed to be the primary drivers of sustainability integration should foster resourceful business practices, be more professionally, culturally, and humanly diverse to reap the most benefits out of a limited amount of resources; this can reduce environmental impact and drive environmental performance, in part by avoiding overt generosity in providing organizational resources on the path towards integration (Rubino & Napoli, 2020).

Besides considering what enables sustainability integration, what are potential inhibitors? Gond, Grubnic, Herzig, and Moon (2012) propose eight constellations between sustainability controls and management controls that are contingent on their degree of mutual integration and whether they are used for diagnostic use (i.e. measurement) or interactively to refine the organization's sustainability strategy through dialogue between top management and lower-

level employees. The framework hinges on the idea of an ideal state of sustainability integration in which traditional and sustainability-oriented control systems are tightly integrated into each other and encourage communication across the organizational hierarchy (ibid.). This state, however, tends to be achieved by transitioning on a path that can have more than one transitional state lying between the assumed ideal and a state in which sustainability is virtually disconnected from the core business and controls, which is called a dormant decoupled strategy (ibid.). For instance, the introduction of sustainability to organizational strategy can either emerge as a strategic need due to voluntary reorientation or can be imposed by regulators requiring compliance with sustainability policies (ibid.). After introducing sustainability to the organization and making its first adoption, the next transitional state tends to be a schizoid strategy characterized through likely deliberate "contrasted sustainable behaviours" (ibid.:213) indicating low sustainability integration, or peripheral sustainability integration as a result of extending compliance to mandated policies to voluntary reporting framework and adopting a long-term operational focus lacking SD engrained in organizational culture that is reflected in routines, usually through an organization-wide monitoring system (ibid.).

Sustainability integration, i.e. the intensity of the interplay between MCSs and SCSs tends to face technological (degree of sophistication in both systems), organizational (capture organizational members' behavior), and cognitive barriers (ability to capture how these members think of sustainability) and become a complex process requiring thorough management (Gond et al., 2012). In a study on the oil and gas sector, George et al. (2016) outline characteristics, enablers and barriers that firms with a disconnected, compliance-driven, or peripheral degree of sustainability integration tend to face.

A dormant decoupled strategy setting tends to be characterized by external pressures and an organization prioritizing economic growth and building organizational learning and capabilities, also on the management level, and control systems, if in place, tend be diagnostic and rather irresponsive to changes in performance (ibid.). Technical barriers comprise underdeveloped controls, lacking formalization in performance evaluation and target setting and isolated HSEQ planning, whereas organizational barriers entail rapid expansion and a lack of structure and formalization for sustainability and related roles and responsibilities (ibid.). The latter tend to be paired with cognitive issues characterized by little industry experience and understanding of higher-level sustainability issues. Regarding enablers, those of technical nature represent a tighter integration of HSEQ into budgeting and auditing processes and the creation of policies for all business units. On the organizational level, HSEQ control should be expanded and in terms of cognitive enablers, the authors recommend increasing awareness on HSEQ issues and relate organizational vision and mission statements to economic and social development issues on the national level (ibid.).

Compliance-driven strategies, on the other hand, also face external pressures yet tend to focus on organizational excellence across processes, publicity, and CSR disclosure reactively (ibid.). MCSs are actively used for feedback, budgeting, and rewarding, and tend to be used interactively, whereas SCSs tend to remain diagnostic and less coherent than the former (ibid.). Technical barriers tend to be characterized by concentrating sustainability planning and performance evaluation in senior management functions and/or a few departments, and limited extent of SR, whereas organizational barriers tend to be characterized by weakly developed implementation of sustainability aspects and related hiring limited to HSEQ (ibid.). From a cognitive perspective, employee focus tends to lie on profits and the roles, responsibilities, and knowledge on the concept and scope of sustainability concerns to the organizations tend to be weakly defined (ibid.). From a technical standpoint, moving beyond compliance can be accomplished by making MCSs and SR more comprehensive and integrating the HSEQ lens across the supply chain (ibid.). On the organizational level, establishing a dedicated ESG framework, focus areas, working groups, and hiring experts tend to be beneficial, whereas cognitive stances can be improved by raising awareness on sustainability issues and obtaining (further) support from top management (ibid.).

Peripheral sustainability integration tends to rest on external pressures, changes in leadership, and/or reputational issues (ibid.). MCSs and SCSs are more tightly integrated and the latter more coherent, most notably by establishing dedicated data collection channels for sustainability data (ibid.). Technically, however, measuring sustainability costs and risks poses a challenge and HSEQ KPIs and environmental impact assessment tend to be limited in their usefulness and the key organizational impediment is limited stakeholder engagement, whereas cognitive issues tend to point at a poorly developed innovation culture and poorly diffused knowledge around sustainability (ibid.). From a technical perspective, this constellation can be enabled through more extensive sustainability planning and linking more HSEQ KPIs to individuals and top managers' compensation schemes, which could (1) potentially be tied to successfully realigning organizational functions, strengthening HSEQ procedures, and capitalizing on sustainability task forces from an organizational standpoint, and (2) coerce changes in top management and employee mindset towards a more holistic understanding of organizational sustainability performance and driving operational impacts (ibid.).

To conclude, transitioning away from decoupled and compliance-driven strategies tends to hinge on cognitive barriers, which emphasize the need to build legitimacy of sustainability concerns and high-powered incentives across organizational levels and departments through such mechanisms (ibid.; Derchi et al., 2020). For instance, forming task forces as groups of promoters and assigning a higher priority to sustainability performance and dual returns, steepening organizational learning curves, and retaining newly gained knowledge and awareness in dedicated MCSs capturing en-

vironmental and social impacts and knowledge repositories (Agrawal & Hockerts, 2019; Figge et al., 2002; Hubbard, 2009; Jackson, 2013; Vermeulen & Barkema, 2002). It should be noted, however, that even when a comprehensive SR, sustainability budgets and efficient incentives are implemented can weak or negative economic performance supersede organizational alignment towards sustainability and prioritize financial profitability over improving nonfinancial performance (Battaglia, Passetti, Bianchi, & Frey, 2016). This implies that sustainability integration can be marginalized despite sophisticated control mechanisms; assuming a beyond-compliance sustainability integration strategy, the adopting organization (here: Cadeler A/S) tends to find itself in a transitional position between a compliance-driven sustainability strategy and peripheral integration of such until it has regained its ability to generate financial returns (ibid.).

2.3.4. Bridging the SR Gap in SMEs

The initial discussion on SR was primarily focused on large-scale corporations which are assumed more likely to have sufficient information-processing capacity to successfully carry out SR and capitalize on the insights provided by a complementary nonfinancial lens (Baron, 2014; Maguire, 2010; Reynolds & Yuthas, 2008). SMEs, smaller in scale yet higher in count, tend to account for most of the business activity, GHG emissions, and employment in the non-public sector yet have witnessed a rather low degree of SR adoption, for which potential reasons are reviewed in this section. A key challenge for such organizations is to "manage their dual mission, integrate social and environmental goals in their business model, and incorporate accountability mechanisms, all while scaling up and garnering the necessary resources to be economically competitive" (Nigri & Baldo, 2018:1). With the GRI Standards as a well-recognized SR framework and IR as an emerging trend in corporate disclosure (Arena & Azzone, 2012; Mauro, Cinquini, Simonini, & Tenucci, 2020), this section focuses on these two alternatives specifically.

Despite seemingly tangible benefits that SMEs can realize trough adopting SR which "are not necessarily less advanced in organizing CSR [Reporting] than large firms" (Baumann-Pauly et al., 2013:693), what prevents them from implementing it? Arena and Azzone (2012) present three impediments to smaller-scale organizations: (1) time constraints, (2) skill gaps in preparing sustainability information due to largely present track records or low sensitivity towards SR along with imprecise expert judgments on resource consumption as a result of such, and (3) low benefits relative to informational cost and managers as business owners having little incentive to move organizational disclosure beyond compliance (Orsato, 2006). Importantly, SMEs tend to be at risk of lacking information-processing capacity to live up to informational requirements set by an increasingly diverse stakeholder audience (Arena & Azzone, 2012). Specifically, empirical evidence on organizing CSR reporting in organizations by Baumann-Pauly et al. (2013) identifies a negative relation between a reporting organization's scale and relative communication cost linked to CSR activity and a positive link between scale and organizational cost of integrating nonfinancial reporting mechanisms into existing routines (ibid.). That is, smaller-scale organizations tend to face lower implementation cost yet much higher reporting cost. The authors describe this discrepancy as a reporting rap, whereas the opposite case holds for larger-scale organizations as they tend to suffer from an implementation gap (ibid.). In the authors' conceptual model, the cost-induced gaps seem to disappear when the organization has obtained a medium scale, which does not mean that SR becomes more cost-efficient but rather that implementation and reporting cost incurred are virtually not prohibitive for either alternative (ibid.).

Lacking information-processing capacity due to low scale economies also tends to be one of the drivers on why reporting organizations initially join such reporting initiatives as the UNGC and become delisted once required to submit a "Communication on Progress" document that discloses an organization's sustainability performance according to the UNGC's principles, which are largely embodied in the NFRD and GRI (European Commission, 2014; Rasche et al., 2020; Shift & Mazars, 2015). Specifically, Rasche et al. (2020) identify a positive link between the decision to remain listed and an organization's scale, early adoption of SR, public listing, and connectedness to a local network promoting SR. The findings imply that late adopters that do not benefit from "legitimacy spillover effects on local networks" (ibid.:1) and join a multi-stakeholder initiative for the sake of certification without the capacity to live up to reporting requirements will face substantially higher challenges to engage in a transparent stakeholder dialogue, provided that the organization's management shows little commitment to such (ibid., Adams and McNicholas (2007)). With Morsing and Thyssen (2003) indicating strong social networks, strong ties between the government, organizations and the media, and partiality to transparency, commitment and fairness in the Danish economic environment, it seems unlikely that Danish UNGC signatories resign as the longer-term opportunity costs would outweigh potential benefits from slack information processing capacity.

Very recent research notes that resource constraints do not only hold with respect to SR, but also to IR. In a series of semi-structured interviews with SMEs' executives, Gerwanski (2020) finds that IR is primarily regarded as a means to improve the organization's image and perceived legitimacy rather than a viable reporting tool that tends to face three major implementation barriers: (1) public addressees do not consider such reports to be relevant or informative, (2) conceptual shortcomings inherent to IR such as a potential overstatement of the importance of shareholders and corporate image, and (3) substantial preparation costs in excess of the surveyed SMEs' reporting capacity. Factoring in that IR tends to be a rather advanced concept that implicitly assumes experience with more traditional SR (Baron, 2014), this study assumes that Cadeler can reap more learning economies by getting started with SR that is holistic yet less integrated than IR and use accumulated reporting knowledge to condense or-

ganizational information into a potentially more condensed and integrated framework in the future.

3. Methodology

3.1. Internal Materiality Assessment

The research method applied in this study attempts to take an outward-looking and an inward-looking perspective on Cadeler's material sustainability topics to effectively live up to the concept of double materiality that tends to emphasize balanced and informative SR and can attract more ESG/impact investors' attention (Agrawal & Hockerts, 2019; Amir & Serafeim, 2018; Quatrini, 2021). It follows the four-step procedure for materiality assessments described in Global Reporting Initiative (2013), in which the identification and prioritization procedures are based on the Fuzzy AHP (FAHP) methodology introduced in Calabrese et al. (2016); Calabrese, Costa, and Menichini (2013). FAHP falls under a broad range of MCDM techniques that have been designed to enable decision-making under uncertainty and consideration of multiple criteria and/or alternatives and tend to be most applicable to complex decision problems with a high impact to the decision that, relative to its consequences, should be better-informed, substantiated, and transparent (Krejčí, 2018). Importantly, these methods intend to not provide an exact correct but rather to support the decision-making process of an individual with imperfect knowledge on his/her own preferences (ibid.). Consequentially, virtually all MCDM methods are bound to entail a certain degree of subjectivity when guiding a DM towards a most preferred rather than an optimal solution (ibid.). A materiality assessment is an inherently subjective process and requires one or more DMs to learn and provide their own preferences on a novel topic to generate a solution that is, to an extent, aligned with their own and external stakeholders' preferences (Beske et al., 2020; Calabrese, Costa, Ghiron, & Menichini, 2017).

3.1.1. Analytical Hierarchy and (Fuzzy) AHP

Fuzzy AHP was initially developed by Van Laarhoven and Pedryczt (1983) as an extension to classical AHP, developed by Thomas Saaty in 1971 and intended to account for more than one analytical objective and deal with potentially unstructured and complex decision problems that involve incomplete information and quantitative and qualitative considerations (Mattiussi, Rosano, & Simeoni, 2014; Wind & Saaty, 1980). Classical AHP decomposes a decision problem into a hierarchical structure of two or more levels and allows a DM to iteratively conduct preference judgments for the next lower level of the hierarchy that contain both subjective and objective considerations in constructing a prioritization (Saaty, 1978). Whereas the procedure is relatively simple to use and tests transitivity of preferences through a maximum eigenvalue approach (ibid.), one of its major drawbacks is that the conversion scale capturing the intensity of a preference consists of crisp numbers that assume that a DM's verbal judgment is perfectly certain and argued to be contradictory to the nature of human thinking (S. Chen & Fan, 2011). Uncertainty can, in such case, evoke different perceptions among individuals such as vagueness that can arise from incomplete information and is similar yet different to the concept of imprecision that points at fuzziness as a moderator of the truthfulness and the value of information used for decision-making (Dubois & Prade, 1988; İbrahim Özkan & Türkşen, 2014). Fuzzy numbers were introduced to remedy this shortcoming and capture this uncertainty and subjectivity inherent to linguistic expressions and have found increasingly broad application within business disciplines (Govindan, Khodaverdi, & Jafarian, 2013; Mardani, Jusoh, & Zavadskas, 2015) and diverse scientific areas (Chan et al., 2019) and are preferred to fuzzifying crisp numbers (Saaty & Tran, 2007). Based on previous assertions, the author converts DMs' judgments into TFNs due to their linearity and simplified arithmetic (Calabrese et al., 2016; Kaufmann & Gupta, 1991; Krejčí, 2018).

To obtain a suggestive GRI-compliant SR and integration agenda for Cadeler, the hierarchy is based on the GRI Standards in effect since 1st July 2018 and expanded by three additional standards on 1st January 2021 (Global Reporting Initiative & University of Stellenbosch Business School, 2020; Matuszyk & Rymkiewicz, 2018) because they tend to be the preferred and most widely dispersed SR framework to adopt (Brown, de Jong, & Levy, 2009; Rodríguez-Gutiérrez et al., 2021) and reflect the impact-oriented multi-stakeholder focus that Cadeler intends to pursue strategically. Finally, Cadeler's former parent company, SPO, uses the GRI Standards, indicating that GRI tends to fit the reporting aspiration of offshore operators. Importantly, the content elements included in the hierarchy are focused on the GRI's reporting requirements and provide a point of departure for GRI-based SR.

The generic method applied in this study, besides its computational and conceptual simplicity, has further advantages over other candidates suggested by Hsu et al. (2013), Bellantuono et al. (2016), and Calabrese et al. (2019). To start with, an analytical hierarchy of the GRI Standards facilitates the preparation of a SR document and renders sustainability performance more comparable across GRI-compliant organizations (Calabrese et al., 2017) allows to set minimum materiality thresholds to quanlitative information in the absence of the five percent rule (Villiers & Maroun, 2017; Morgan, 2014; Whitehead, 2017). In addition, relative priorities (weights) obtained in the process are additive and allow a reporting organization to set a coverage level to economize on organizational resources and expertise, render the most material content elements as informative as possible and aggregate fewer material items into summary descriptions; this quantitative tracker is largely absent in alternative frameworks (ibid.). The ability to aggregate stakeholder judgment is consistent with the GRI Standard' multi-stakeholder nature and a step-wise procedure mirrors the GRI's recommendation in that they recommend to first prioritize material topics and disclosures (i.e. sub-topics) in that order (Global Report-

ing Initiative & University of Stellenbosch Business School, 2020). Given time constraints in the data collection process, this approach was selected as it can be used to factor in external stakeholders' considerations through an internal perspective (Calabrese et al., 2016) which the author instructed survey participants to do. This procedure can potentially reduce endogeneity bias and is consistent with the idea that MCDM techniques intend to provide decision support rather than a definitive solution (Krejčí, 2018).

The approach used for this study entails a few modifications compared to Calabrese et al. (2016). First, it can test the results for their robustness by changing DM bias and the nature (COA) of the TFNs (C. W. Chang, Wu, Lin, & Chen, 2007; Tsai, Chang, & Lin, 2010). Second, the method is based on a more recent, comprehensive, and mandating set of GRI ruling. Third, a tolerance mechanisms for seemingly inconsistent preferences is explicitly provided (Alonso & Lamata, 2006). The approach aggregates input data from several specialists across Cadeler, following Herremans and Nazari (2016) in that the prioritizing and reporting on material sustainability issues is the aggregate of specialist preferences throughout the organization to obtain sufficiently objectified priorities of Cadeler A/S as a single collective DM (ibid.; Aull-Hyde, Erdogan, & Duke, 2006).

The materiality survey consists of nine questionnaires covering preferences on subcategories, topics, and disclosures across the GRI's economic, environmental, and social standards. Though social subcategories are absent in the GRI Standards, the author added a subcategory layer known from the GRI G4 (Global Reporting Initiative, 2013) and assigned standards to each subcategory by mapping them to their G4 "aspects" equivalent through a mapping provided by the GRI to ease the transition from the G4 to the Standards (Global Reporting Initiative, 2017). Doing so reduces computational effort and decision fatigue for the author and surveyed DMs as the number of pairwise comparisons decreases substantially (Calabrese et al., 2017, 2016). An additional subcategory ranking intends to mimic the GRI Standards' modular structure and topics' sensitivity.[1]

Because sustainability performance is driven by inherently quanlitative factors and entails substantial multidimensionality and potential interrelatedness (Venturelli et al., 2020), integrating the GRI Standards into the Fuzzy AHP framework reduces complexity by a clear distinction between sustainability (sub-)categories, topics, and disclosures (Calabrese et al., 2016). In addition, each GRI disclosure is only connected to one higher-level topic, which is connected to a (sub-)category higher up in the analytical hierarchy (ibid.). To illustrate, the framework's structure distinguishing between economic, environmental and social categories is intended to (1) facilitate prioritization of content elements by proceeding from topics to disclosures, (2) put equal emphasis on each of these categories to provide a balanced picture on organizational sustainability performance sustainability performance, and (3) signal credible contribution towards SD and aspiration levels within each of these categories by encouraging decision-relevant SR (Calabrese et al., 2016; Global Reporting Initiative, 2020a; Robinson, 2004; Torelli et al., 2020).

Similar to the framework proposed in previous works, the hierarchy of the modular GRI Standards expands over four levels: Goal, Criteria, Sub-Criteria, and Alternatives (Calabrese et al., 2016, 2013). The goal lies in the prioritization of material GRI topics and disclosures to Cadeler A/S to plan the SR process more efficiently. The criteria represent the economic, environmental, and social perspectives on sustainability that are embodied by the Standards' categories. The third (sub-category) layer lists the economic and environmental topics and social subcategories and topics that provide guidance on which content elements should be paid most attention to. The fourth and lowest level contains management approach and topic-specific disclosures that are only linked to one higher-level sub-criterion.

3.1.2. Translating DM Judgment into SR Priorities

From the collection and compilation of DMs' preferences to the prioritization of economic, environmental, and social sustainability topics and disclosures, there are four steps involved in the framework. In large parts, the analytical procedure is based on the Fuzzy AHP approach proposed by Calabrese et al. (2016, 2013) yet provides three additions to the approach: (1) a tolerance measure that further allows a resource-constrained reporting organization to vary the degree of consistency to avoid resubmissions of preferences that may incur additional informational cost (Alonso & Lamata, 2006), an updated hierarchy that allows the reporting organization to prioritize the latest set of SR content elements provided by GRI, and (3) a sensitivity mechanism to also validate the internal results with respect to changes in DM attitude and degrees of uncertainty exhibited in decision-making (Balusa & Gorai, 2019).

In the first step of the analysis, a decision maker's verbal preferences are compiled in a square PCM $\tilde{A}$ (1):

$$\tilde{A} = \left(a_{ij}^{\alpha}\right)_{n\times n} = \begin{bmatrix} (1,1,1) & \dots & (l_{12}^{\alpha}, m_{12}^{\alpha}, u_{12}^{\alpha}) & \dots & (l_{1n}^{\alpha}, m_{1n}^{\alpha}, u_{1n}^{\alpha}) \\ (l_{21}^{\alpha}, m_{21}^{\alpha}, u_{21}^{\alpha}) & \dots & (1,1,1) & \dots & (l_{2n}^{\alpha}, m_{2n}^{\alpha}, u_{2n}^{\alpha}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ (l_{n1}^{\alpha}, m_{n1}^{\alpha}, u_{n1}^{\alpha}) & \dots & (l_{n2}^{\alpha}, m_{n2}^{\alpha}, u_{n2}^{\alpha}) & \dots & (1,1,1) \end{bmatrix} \quad (1)$$

Where:

$$a_{ij}^{\alpha} = \left(l_{ij}^{\alpha}, m_{ij}^{\alpha}, u_{ij}^{\alpha}\right) = \left(a_{ji}^{\alpha}\right)^{-1} = \left(\frac{1}{u_{ij}^{\alpha}}, \frac{1}{m_{ij}^{\alpha}}, \frac{1}{l_{ij}^{\alpha}}\right) \quad i,j = 1,\dots n; i \neq j \quad (2)$$

represents a fuzzy number for the linguistic judgment(s) a decision maker assigns to an alternative i relative to alternative j on a nine-point triangular fuzzy conversion scale (D. Y. Chang, 1996; S. H. Lee, 2010).

[1] See table 1 in the appendix.

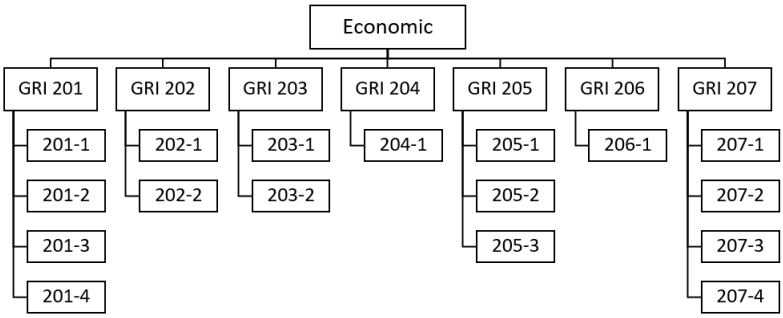


Figure 1: Hierarchy of Economic Topics and Disclosures (own work)

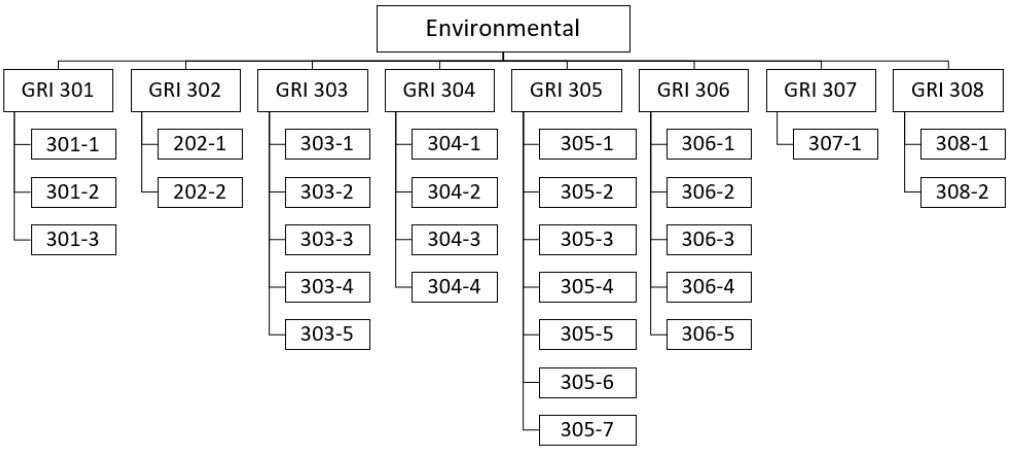


Figure 2: Hierarchy of Environmental Topics and Disclosures (own work)

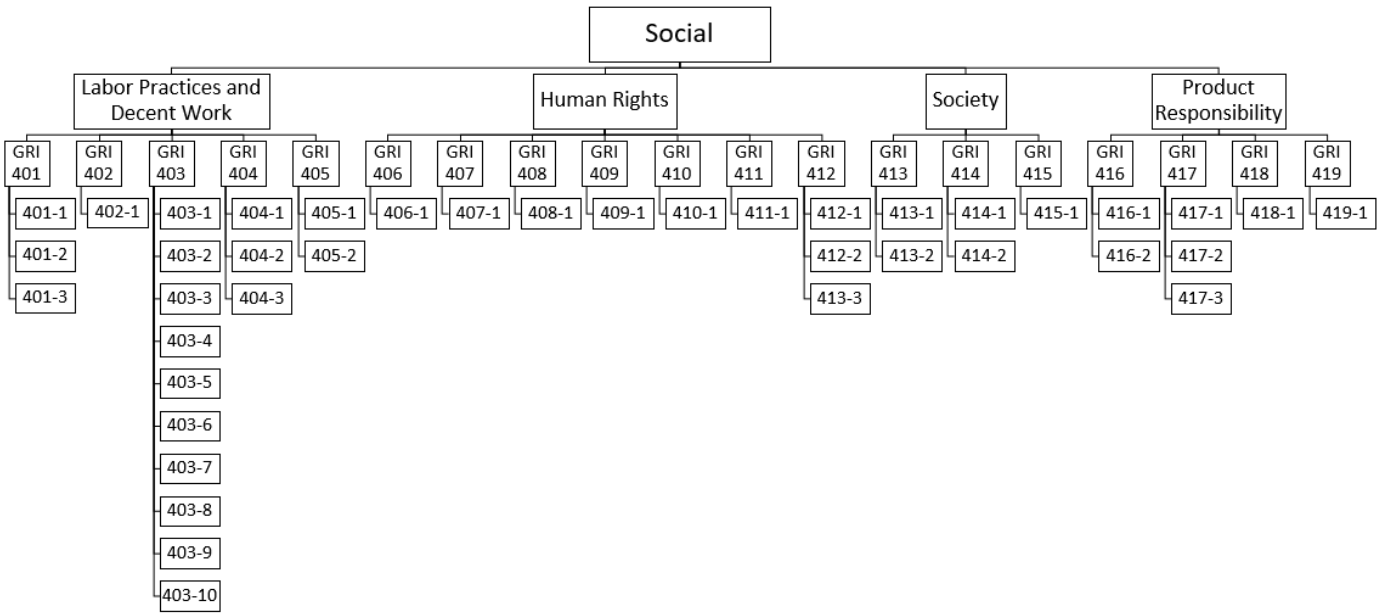


Figure 3: Hierarchy of Social Subcategories, Topics, and Disclosures (own work)

A TFN a_{ij}^{α} has the following membership function (Kaufmann & Gupta, 1991):

$$a_{ij}^{\alpha}(x) = \begin{cases} \frac{x-l}{m-l}, & l \le x \le m \\ \frac{u-x}{u-m}, & m \le x \le u \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Where m has the highest extent of $a_{ij}^{\alpha}(x)$ and l and u represent the TFN's upper and lower bounds that can overlap depending on the level of fuzziness α inherent to the TFNs (Balusa & Gorai, 2019). Figure 4 displays the TFNs for the analysis based on Table 2 in the appendix.

The method is based on the Fuzzy AHP method proposed by Calabrese et al. (2016, 2013) to identify the relative weights of the GRI subcategories, topics, and disclosures without assigning illogical zero weights to DM preference sets (ibid.) and expands it by a generic defuzzification mechanism that allows for robustness tests given that rank reversal tends to be a common problem in AHP-based approaches (Krejčí, 2018). When compared to traditional fuzzy extensions of classical AHP, the fuzzy extension of the approach applied in this study is limited to capturing verbal uncertainty in the DMs' preferences (Calabrese et al., 2016). Primarily, this setup is intended to reduce computational complexity to a considerable extent and preserve the method's relative simplicity through the calculation of crisp eigenvalues (Krejčí, 2018).

The content elements' materiality scores are determined in a four-step process. In Step 1, a DM's fuzzy PCM defined in equation (1) is converted into a crisp PCM with the centroid or rather center of area (COA) method (Takagi & Sugeno, 1985). This approach increases computational simplicity and tends to yield "well accepted results" (Krejčí, 2018:33). In its generic form and this study, TFNs are defuzzified with a weighted average proposed by A. R. Lee (1995) that captures a DM's bias and uncertainty inherent to the preference set through equation (4):

$$\tilde{a}_{ij}^{\alpha} = \lambda u_{ij}^{\alpha} + (1-\lambda) l_{ij}^{\alpha}, i, j = 1, \ldots, n \quad (4)$$

Where u_{ij}^{α} and l_{ij}^{α} denote the upper and lower bound of a triangular fuzzy number defined in a DM's PCM. λ captures the condition of the DM's attitude and can take the values 0, 0.5 and 1 to model pessimistic, neutral, or optimistic conditions, respectively (Balusa & Gorai, 2019; A. R. Lee, 1995). α, on the other hand, acts as a proxy for fuzziness, i.e. the degree of vagueness under which the preferences are set (ibid.) and ranges from 0 to 1 in increments of 0.2. In the base case analysis, the author assumes that all DMs (1) are unbiased in setting a preference towards and alternative and (2) face a medium level of uncertainty in their decision-making, both of which are indicated by $\lambda = \alpha = 0.5$ and in line with the TFN conversion scale suggested by D. Y. Chang (1996) and S. H. Lee (2010). In addition to the benefit of testing (aggregate) priorities for their robustness, this modification is assumed to be more appropriate for mimicking conditions in which different levels of DM bias and sentiment towards preferences are present. Compared to the defuzzification formula proposed by Wang and Elhag (2007) and used originally by Calabrese et al. (2013), the preferences for the base case scenario remain qualitatively unchanged.

In Step 2, the PCMs' consistency is analyzed with the Maximum Eigenvalue Method proposed by Saaty (1978). A PCM's consistency and preferences' transitivity are measured through a CI (6) and CR (7):

$$CI = \frac{(\lambda_{\max} - n)}{n-1} \quad (6)$$

$$CR = \frac{CI}{RI} \quad (7)$$

Where $\lambda_{\max}$ is a PCM's largest eigenvalue, n its dimensionality and RI a random index contingent on the PCM's dimensionality which is a "CI expected from a matrix of that order [dimensionality]" (Donegan & Dodd, 1991:135).

Conventionally, preferences tend to be consistent (transitive) when the CR of a PCM is lower than 10%. Although traditional MCDM research suggests that a DM should reevaluate his or her preferences with a new PCM in the case of inconsistency, Alonso and Lamata (2006) assert that this threshold can be adjusted relative to the DMs' tolerance towards (in)transitivity/consistency or a larger sample of decision makers' preferences that are aggregated to infer possible collective priorities. This mechanism also provides DMs with a certain degree of flexibility and compensation for decision fatigue, for instance when DMs tend to be less familiar with the survey set which may be detrimental to transitive preferences; in addition, it allows a reporting organization to further economize on SR-related resource commitment (ibid.).[2]

Third, local priority weights for each sub-category, topic, and/or disclosure item are computed by taking the row sums of a consistent fuzzy PCM $\tilde{A}$ (8) and then normalizing them with equation (9) to obtain $\tilde{S}_j$ (Wang, Luo, & Hua, 2008). The latter enables a more precise normalization of fuzzy weights as it considers the fuzzy weights' interdependence (see Wang & Elhag, 2006 for its derivation and a detailed discsussion):

$$\widetilde{RS}_i^{\alpha} = \sum_{j=1}^{n} \tilde{a}_{ij}^{\alpha} = \left(\sum_{j=1}^{n} l_{ij}^{\alpha}, \sum_{j=1}^{n} m_{ij}^{\alpha}, \sum_{j=1}^{n} u_{ij}^{\alpha} \right), \quad i = 1, \ldots, n \quad (8)$$

$$\begin{aligned} \tilde{S}_i^{\alpha} &= \frac{\widetilde{RS}_i^{\alpha}}{\sum_{j=1}^{n} \widetilde{RS}_j^{\alpha}} \\ &= \left(\frac{\sum_{j=1}^{n} l_{ij}^{\alpha}}{\sum_{j=1}^{n} l_{ij}^{\alpha} + \sum_{k=1, k\neq i}^{n} \sum_{j=1}^{n} u_{kj}^{\alpha}}, \frac{\sum_{j=1}^{n} m_{ij}^{\alpha}}{\sum_{k=1}^{n} \sum_{j=1}^{n} m_{kj}^{\alpha}}, \right. \\ &\quad \left. \frac{\sum_{j=1}^{n} u_{ij}^{\alpha}}{\sum_{j=1}^{n} u_{ij}^{\alpha} + \sum_{k=1, k\neq i}^{n} \sum_{j=1}^{n} l_{kj}^{\alpha}} \right) \\ &= \left(l_i^{\alpha}, m_i^{\alpha}, u_i^{\alpha} \right), \quad i = 1, \ldots, n \end{aligned} \quad (9)$$

[2] See table 3 and 4 in the appendix.

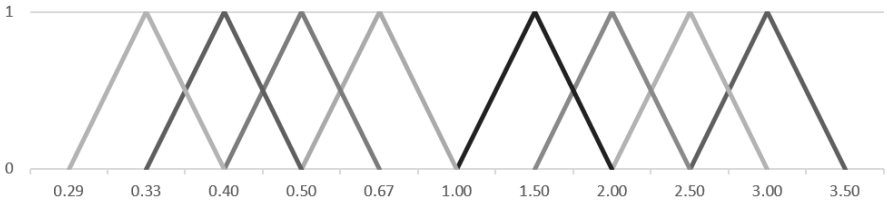


Figure 4: TFNs' Membership Functions for $\alpha = 0.5$ (based on D. Y. Chang, 1996; S. H. Lee, 2010)

Finally, the crisp numbers obtained through equation (10) are normalized to obtain the DM's local materiality vector for all local alternatives in the hierarchy (Balusa & Gorai, 2019; Calabrese et al., 2016).

$$w_i^{\alpha} = S_i^{\alpha}\left(\tilde{S}_i^{\alpha}\right) = \lambda u_i^{\alpha} + (1-\lambda) l_i^{\alpha}, i, j = 1, \ldots, n \tag{10}$$

$$W^{\alpha} = \left(w_1^{\alpha\prime}, w_2^{\alpha\prime}, \ldots, w_n^{\alpha\prime}\right) \tag{11}$$

Since the economic, environmental, and social surveys are completed by multiple DMs in Cadeler A/S, the individual preferences are synthesized into a single aggregate PCM per category, sub-category, and/or topic. Assuming (1) a similar level of expertise on the GRI Standards and (2) a similar degree of perceived novelty among the surveyed DMs along with high similarity in ranking preferences under the consideration of both Cadeler's interests and the organization's external stakeholders' informational needs each surveyed individual in Cadeler A/S was assigned the same level of importance in the aggregation process. Furthermore, this approach tends to be representative of the participatory decision-making in organizing SR within Cadeler A/S.

If formula (12) denotes the TFN provided by a decision maker $k(k = 1, \ldots m)$ and (13) its reciprocal, then (14) represents the average judgment or rather collective preference according to the fuzzy addition of TFNs (Kaufmann & Gupta, 1991). In the model of Calabrese et al. (2016), the aggregate PCM obtained through (14) is then used to infer the collective preference set by utilizing the above four-step approach.

$$\tilde{a}_{ij}^{(\alpha k)} = \left(l_{ij}^{(\alpha k)}, m_{ij}^{(\alpha k)}, u_{ij}^{(\alpha k)}\right) \tag{12}$$

$$\left(\tilde{a}_{ij}^{(\alpha k)}\right)^{-1} = \left(\frac{1}{u_{ij}^{(\alpha k)}}, \frac{1}{m_{ij}^{(\alpha k)}}, \frac{1}{l_{ij}^{(\alpha k)}}\right) \tag{13}$$

$$\overline{a_{ij}^{\alpha}} = \frac{1}{m}\sum_{k=1}^{m} \tilde{a}_{ij}^{(\alpha k)} = \left(\frac{1}{m}\sum_{k=1}^{m} l_{ij}^{(\alpha k)}, \frac{1}{m}\sum_{k=1}^{m} m_{ij}^{(\alpha k)}, \frac{1}{m}\sum_{k=1}^{m} u_{ij}^{(\alpha k)}\right) \tag{14}$$

Given the relative small range in the verbal conversion scale, an arithmetic mean tends to yield qualitatively similar conclusions to a geometric mean and its application to all a_{ij}^{α} only preserves reciprocity in quantification (Aczél & Saaty, 1983; Aull-Hyde et al., 2006; Pandey, 2012). In addition, the DMs' judgments were given equal importance in the analytical model for three reasons. First, adopting a structured materiality assessment is a novelty to the organization and assumes a similar extent to which the survey participants are familiar with the implications of a quantitative approach. Second, the surveyed individuals have a similar degree of influence on which content elements are deemed more material and/or disclosed externally to Cadeler's cross-industry stakeholder audience. Finally, equal weighting accounts for alike decision-making power on the scope of organizational processes that are assumed to be linked to the content of the preference surveys.

3.1.3. Application to Cadeler A/S

Cadeler A/S operates WIVs primarily used for the transportation, installation, and maintenance of offshore wind turbines and their foundations. As such, the company's operations are a critical link in the offshore wind supply chain enabling the transition of the electrical grid to renewable energy sources. The organization is certified under ISO 9001:2015 and ISO 14001:2015 and strictly follows ISO 26000:2015 which signals strong organizational commitment to SD across economic, environmental, and social areas such as long-term value creation, biodiversity preservation and circular economy, and the promotion of diversity and empowering employees. Their four-pronged SD strategy is summarized in Figure 5.

The scope of potentially material topics and disclosures was identified through semi-structured interviews with the organization's sustainability and environment manager. Despite the possibility to aggregate immaterial or non-applicable items into summary positions, all GRI Standards were deemed applicable and included in the internal materiality assessment. Internal data for the internal FAHP study was collected from generalist and specialist managers across such functions as Finance, Operations, HSEQ, Sustainability, Procurement, Tender Management, Marine Operations, and General Management. The DMs tend to be most knowledgeable on Cadeler's and their stakeholders' preferences based on frequent exposure through project-level cooperation and correspondence and can consider the latter in completing the PCMs they are assigned to. Out of ten DMs contacted, the survey realized a response rate of 50%. Aggregation ensures the respondents' anonymity given the small sample and scale

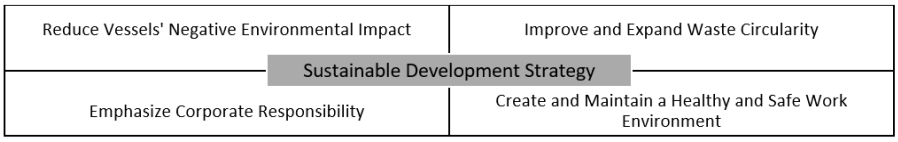


Figure 5: Cadeler A/S SD Strategy (Cadeler A/S, 2021)

of the organization, simulates group-level decision-making and further ensures a democratic balanced perspective on content elements' relative importance and a snapshot of the organization's SR priorities (Aull-Hyde et al., 2006).

The surveys were tailored to generalist and specialist managers in that the former received surveys on the full scope of GRI subcategories and topics only, whereas the latter were invited to provide preferences on disclosures in addition. This structure is consistent with the idea that specialist knowledge is embedded in such departments and that consistent yet non-indifferent preferences are provided as a result of higher topic familiarity (Schläpfer & Fischoff, 2010). Preference data was collected through Excel-based surveys over a period of four weeks to account for time constraints and to increase the likelihood of obtaining consistent preferences (CR $\leq$ 10%). Each survey included a completion manual and definitions on the GRI Topics and Disclosures; in addition, the author provided tutorials and content element definitions in the survey material and organized online walkthroughs to ensure DMs' understanding of the survey's objective, the mechanisms in play, and how DMs can contribute to their best knowledge. The surveyed DMs were familiarized with the content elements' hierarchy and are assumed to have set clear preferences on which subcategories should be prioritized over others to assign, ceteris paribus, a higher importance to the social topics assigned to the subcategory. For illustration, the first pairwise comparison in Table 6 in the appendix answers the following question:

> "How much more (or less) informative is the GRI Topic "Materials" compared to "Energy" to accurately describe Cadeler's activities' environmental impact and sustainability performance?"

Following the same architecture of the question for topics and disclosures, the respondents set preferences and were asked to reflect the content elements' materiality in terms of strategic importance rather than their ease of reporting or related data collection (Puroila & Mäkelä, 2019). Prior to aggregating individual DMs' preferences into a collective preference set for Cadeler, each submitted PCM was tested for consistency and transitive preferences. Untabulated results indicate that all submitted PCM exhibit transitivity (CR $<$ 10%) and are aggregated for the base case analysis, of which the results also point at transitivity. Across all GRI Standards and DMs, the author collected 492 pairwise comparisons.[3]

[3] See tables 5, 6 and 7 in the appendix.

3.2. External Materiality Assessment

Due to research and temporal constraints arising from reputational and strategic concerns that inquiries to external stakeholders may result in reporting requirements that may be misaligned with Cadeler's priorities, prioritizations from this stakeholder group is not obtained but rather derived from the materiality sections of SRs from organizations that Cadeler A/S is frequently exposed to due to the nature of their business (Machado et al., 2021). Cadeler's stakeholders can be segmented into three major groups, namely (1) Direct Competitors, (2) OEMs, i.e. offshore wind turbine manufacturers, and (3) Windfarm Developers. A comprehensive overview on 19 stakeholders based on the organization's track record has been derived from the organization's website and extended by records provided by Cadeler's sustainability management. Though Cadeler's NACE Rev 2 Code according to Orbis (Bureau van Dijk) is 50.20 ("Sea and coastal freight water transport"), setting up a peer group based on this code yields organizations that are not occupied with offshore windfarm commissioning. Therefore, Cadeler's primary economic activity is more accurately described through NACE Code 42.91 ("Construction of Water Projects") since Orbis classifies direct competitors as such. Furthermore, the competitors' revenue streams arising from offshore windfarm commission on a vessel basis are deemed comparable to Cadeler's. OEMs are characterized by the NACE Code 28.11 (Manufacturing of wind turbines) whereas Windfarm developers exhibit the NACE Code 35.11 (Production of electricity).

This type of manual content analysis is beneficial in that the absence of direct inquiries to external stakeholders is assumed to avoid the short-term emergence of inapplicable reporting requirements to Cadeler A/S. On the other hand, it assumes that topic priorities are truthfully reported and informative on the organization's true preferences and not constructed or tailored towards a specific stakeholder group (Mitchell et al., 1997). Due to lacking standardization in the SRs' topic descriptions, the author mapped each topic included in an organization's materiality matrix to its closest GRI equivalent based on the GRI's Standards catalogue and available information in the stakeholders' reporting material. The analysis was restricted to the topic level to avoid overinterpreting content from potentially non-GRI compliant reports and to focus on higher-level priority alignment. Some organizations reported against SDG targets which required translation though a dedicated GRI-SDG mapping (Global Reporting Initiative, 2020b). Broadly speaking, the more recurrent a GRI topic is, the more material it is assumed (Bellantuono et al., 2016). Topic rankings were constructed

based on their recurrence. Within stakeholder groups, the materiality sections were, on average, similarly comprehensive among direct competitors, OEMs, and windfarm developers reporting on 17, 16, and 20 GRI topics, respectively. Equal stakeholder salience within and across groups is assumed and stakeholder groups are analyzed independently to avoid oversimplification of priorities (Puroila & Mäkelä, 2019). Spearman rank correlations and two-tailed significance tests examine the derived ordinal data's explanatory power (Gauthier, 2001).

3.3. SDG Prioritization

Performance reporting against the UN's SDGs is becoming increasingly important for tracking progress towards a carbon-neutral economy by 2030 (Adams, 2017; KPMG, 2020; United Nations, 2015b). Notably, Adams (2017) emphasizes that reporting organizations should (1) align the SDGs with their sustainability strategy and (2) identify environmental issues impeding shared value creation and therefore contribution and progress towards SD and growth (Robinson, 2004). Implying that the UN can be perceived as a Green Club can strengthen organizational legitimacy to a degree similar to that the GRI can realize, the SDGs represent a convenient solution for reporting on progress towards SD (Adams, 2017; Orsato, 2009).

Out of the few studies that explore and call for more coverage on the link between policy abidance and improved sustainability performance over time (Conway, 2019), very recent developments identify a significant positive association between SDG-driven reporting and progress towards them and the adoption of the GRI Standards (Pizzi et al., 2021). Therefore, Cadeler could enhance their SR's legitimacy, credibility and value relevance by tightly integrating the SDGs and potentially reap benefits from improved stakeholder engagement and recalibrating their SD(G) strategy (Adams, 2017; Adams & Larrinaga, 2019). In the same vein, KPMG (2020) calls for more extensive SDG integration, stronger strategic alignment, greater emphasis on biodiversity. A recent joint project by the GRI, the PRI and the UNGC has resulted in extensive guidance for reporting organizations in how to address investors' informational needs on SDG performance by reporting on the GRI Standards (Global Reporting Initiative, 2020b). The author applies the suggested mapping to the priorities obtained from the internal materiality assessment to display how Cadeler could report on the SDGs, which increasingly shape SR policymaking in the EU (Global Reporting Initiative, 2020a). In the tables with the results, proposed SDG and/or target substitutes (complements) are indicated in brackets and separated without (with) a comma from the GRI's official translation.

4. Results

4.1. Preliminary Insights on Cadeler A/S

Regarding the internal materiality assessment, all DMs' and 38 aggregate PCMs were found to be consistent (CR < 0.1) (Aull-Hyde et al., 2006; Saaty, 1978). Internal results are complemented by their S.D. and CV to proxy dispersion among DMs. Topics' and disclosures' descriptions are taken from the most recent GRI Standards catalogue from 1st January 2021 (Global Reporting Initiative & University of Stellenbosch Business School, 2020).

Concerning Cadeler's present degree of sustainability integration, MCSs are used interactively, whereas SCSs are primarily used diagnostically yet are gaining increased feedback. Unstructured management interviews revealed that compliance is critical to ensure operational excellence and profitability. Regarding environmental KPIs, Cadeler's SCS captures data for activities onshore and offshore. For the former, flight data, fuel consumption of company cars, electrical consumption and waste production are captured. On vessels, MGO consumption, flight data, waste production, paper, and plastic recycling rates along with freshwater consumption, bunkering data, and use and spills of chemicals or hazardous materials along with reported incidents and pollution events (e.g. through ballast water). In terms of employment, Cadeler A/S captures such data as staff diversity, retention rates, and salary and benefit benchmarks on the industry level. The organization's HSEQ system tracks lost time incidents relative to manhours worked, the number of vessel track observations, near misses and hazard IDs to indicate the focus level on onboard safety. Despite a rather comprehensive baseline focus on nonfinancial performance, economic profitability tends to be prioritized which may be detrimental to sustainability integration (George et al., 2016). However, recent initiatives to intertwine administrative processes, strengthen HSEQ, and plans to tie remuneration more closely to sustainability performance indicate a proactive movement beyond compliance and a transition towards peripheral sustainability integration, mainly because of ongoing process innovation, departmental integration, and sustainability knowledge diffusion (ibid.).

4.2. Internal Materiality Assessment

4.2.1. Economic Priorities

Figure 6, Figure 7, and Table 8 in the appendix show the global weights for economic topics and disclosures for Cadeler A/S which represent the content elements' relative importance on the topic and disclosure level and for the organization to report their GRI-based economic sustainability performance internally and to their stakeholders along with their dispersion scores.[4]

GRI 201 (Economic Performance) was given the highest materiality score of 18.46% on the organizational level. It also displays the highest standard deviation across all seven alternatives at 6.21% yet the second-highest relative variation at a CV of 0.34. The topics with the second-and third-highest priority are Anti-Corruption (GRI 205), and Anti-Corruptive Behavior (GRI 206), with materiality scores of 16.82%, and 16.29% and similar CVs of 0.34 and 0.33. Tax

[4] See tables 1 and 8 in the appendix.

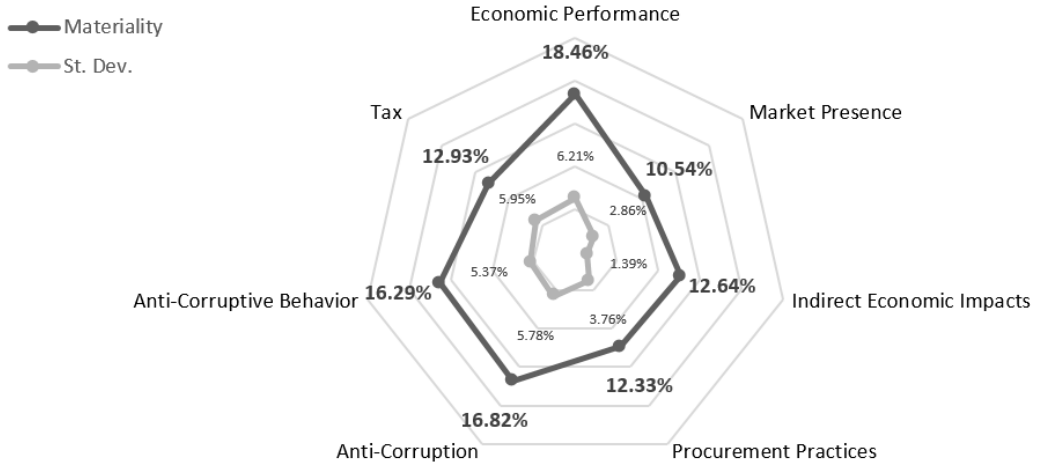


Figure 6: Materiality and Dispersion of Economic GRI Topics (own work)

(GRI 207) is ranked fourth at 12.93% yet has the highest CV in the list (0.46). Indirect Economic Impacts (GRI 203) and Procurement Practices (GRI 204) assumed materiality (CVs) of 12.64% and 12.33% (0.11 and 0.30) along with GRI 207, displaying a similar level of relative importance on the organizational level yet lower absolute and relative variability. Market Presence (GRI 202) is ranked lowest at 10.54% and entails the second-lowest CV across economic topics at 0.27. Economic, environmental, and social topics have an average CV of 0.31, 0.19, and 0.32, respectively. Thus, GRI 302's CV is below-average across economic and social topics yet above-average relative to environmental content elements. All economic topics' CVs except for Tax are about average. To the author's surprise, individual submissions went against the author's expectation that these reflect a DM's task familiarity (Schläpfer & Fischoff, 2010). To illustrate, when a strong bias to procurement practices was expected, the emphasis lied on coverage on anti-corruption and anti-corruptive behavior.

Figure 7 depicts the global weights of Economic GRI disclosures and their contribution towards a report that assumes full coverage on all materiality prioritization of the economic GRI Disclosures for the Case of Cadeler A/S. It further conveys within-topic priorities should Cadeler decide to cover all topics to an extent. Within GRI 201, the most material disclosures are GRI 201-2 and GRI 201-1 at global weights of 8.28% and 6.08% that deal with financial implications, opportunities, and risks related to climate change and the generation and distribution of economic value. Regarding GRI 205, disclosure 205-3 was assigned the highest priority and covers the quantity of and actions taken against corruption incidences on the organizational, employee, supplier, and public level. GRI 206 ranked third only includes GRI 206-1, which reports on exposure to legal action because of anti-competitive behavior and the outcome of the trials and of which the disclosure weight equals that of the topic and represents the most material disclosure item across the economic scope. With regards to Tax, GRI 207-1 to GRI 207-3 represent management approach disclosures that need to be reported in case GRI 207 is adopted. The results show that these disclosures related to the organization's tax-related approach, governance, and stakeholder engagement are ranked highest in this order, whereas disclosure on country-by-country reporting is ranked lowest within GRI 207 and across economic disclosures. GRI 203, GRI 204, and GRI 202 are covered most effectively by reporting on GRI 203-1 and GRI 203-2, GRI 204-1, and GRI 201-1 or 202-2. None of these topics contains management approach disclosures (see Global Reporting Initiative & University of Stellenbosch Business School, 2020).

4.2.2. Environmental Priorities

Figure 8, Figure 9, and Table 9 in the appendix display the relative materiality of environmental topics and disclosures across and within content elements to capture Cadeler's sustainability performance in this category, respectively, along with dispersion scores. At a materiality score (CV) of 22.87% (0.19), Environmental Compliance (GRI 307) is ranked as the most material topic at medium relative variation ahead of Energy (GRI 302) with a considerable margin of 7.51 percentage points. In addition, its relative CV is about average (0.19). Energy (GRI 302) is ranked second at 15.36% and a CV of 0.08, the lowest among environmental topics. Emissions (GRI 305), Waste (GRI 306) and Water and Effluents (GRI 303) complete the list of the top five environmental topics to achieve a GRI coverage of 75.43% achieved by reporting on all 25 disclosures related to the topics. The top three topics and their 13 disclosures would cover 50.43% of the environmental GRI scope. Assuming linear economies of scale in preparing reporting internal and external environmental SR for each topic and disclosure item, reporting costs tend to increase relative to the level of comprehensiveness achieved in the research process. Interestingly, Biodiversity (GRI 304) and Materials (GRI 301) are considered least material (8.22% and 8.03%) yet entail the highest CVs of 0.31

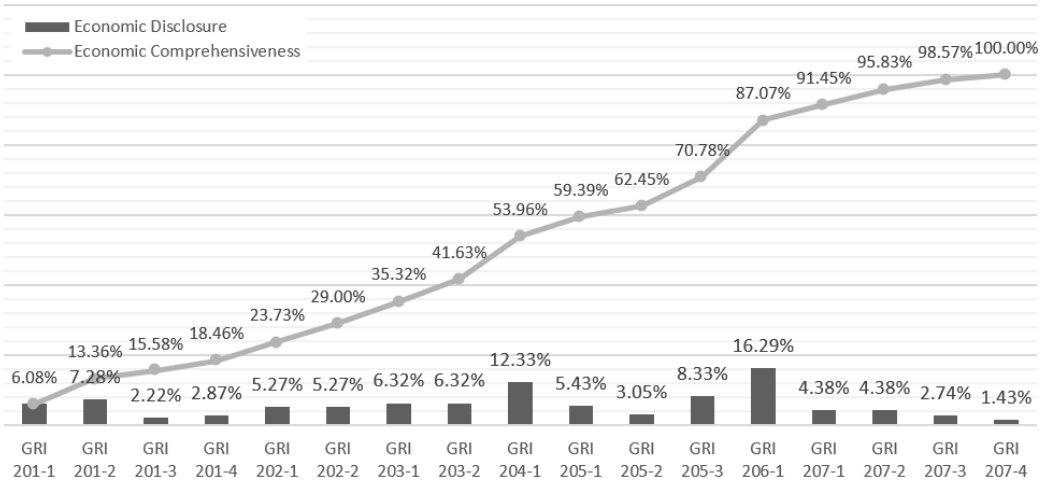


Figure 7: Comprehensiveness and Materiality of Economic GRI Disclosures (own work)

each.[5]

Figure 10 describes environmental disclosures' individual contribution to covering the required content elements included in the environmental GRI Standards. Starting with Environmental Compliance (GRI 307), its only disclosure 307-1 has the same materiality score and thus has the highest individual contribution within environmental disclosures but also across all 89 subtopics included in the framework. Within Energy (GRI 302), over 75% of the topic is covered by the disclosures GRI 302-3, 302-4, and 302-1 that require reporting on Energy intensity (i.e. relative efficiency), reducing energy consumption and energy consumed within Cadeler A/S, respectively at local (global) scores of 28.09%, 24.29%, and 22.65% (4.32%, 3.73%, and 3.48%). Concerning Emissions (GRI 305), GHG Emissions intensity (305-4), Actions to reduce GHG emissions (GRI 305-5), and Direct (Scope 1) emissions (GRI 305-1) are ranked highest at global weights of 3.26%, 2.57%, and 2.46%, and are over 50% informative on Cadeler's environmental performance on Emissions. With regards to Waste (GRI 306), most priority was assigned to the disclosures GRI 306-3 and 306-4 which shed light on how much (hazardous) waste Cadeler A/S has generated and/or diverted from disposal (i.e., reused or recycled) along with contextual information on how data was compiled and segmented according to internal definitions. Notably, GRI 306-1 (waste-related impacts in the organization's value chain) and 306-2 (actions to monitor and mitigate waste-related impacts) were not ranked highest yet constitute mandatory disclosure elements as per Global Reporting Initiative and University of Stellenbosch Business School (2020).

4.2.3. Social Priorities

Figure 10, Figure 11, and Table 10 in the appendix describe the relative importance of social topics to Cadeler A/S on the topic level. In an attempt to update the analytical framework with the latest generation of GRI reporting elements and balance the framework's modularity and the resulting higher cognitive effort and potential decision fatigue and inconsistent preferences when setting preferences on social reporting topics, the author added a social subcategory layer from the GRI G4 (Calabrese et al., 2016; Global Reporting Initiative, 2013) and determined preferences on a subcategory level to mimic a topic ranking absent subcategories.[6]

Figure 10 shows the global weights of social sustainability topics, which are calculated as the product of the of their local weight within a subcategory and the that of the subcategory they are part. Occupational Health & Safety (GRI 403), the highest-ranking topic among all social topics and highest-ranking in the LPDW subcategory, has a global weight of 8.43% = 27.75% * 30.37%. GRI 403 also faces the highest CV of 0.63, which is almost twice the average CV of social topics. The score suggests that Cadeler should commit a maximum of 8.43% of resources committed to social SR to the disclosure elements contained in GRI 403. Within LPDW, Employment (GRI 401) is ranked second (fifth) locally (globally) at 23.33% (7.08%) with a CV of 0.39 and covers hiring and firing, and benefits to full-time employees in office or on (parental) leave.

GRI 416 (Customer Health and Safety) is ranked second across all social topics and reports on an organization's commitment and performance to/on increasing service safety and ensuring such by covering incidents that are a result of non-compliance to safety guidelines. Its materiality (CV) amounts to 7.08% (0.38). Notably, GRI 416 is deemed most material within the PR subcategory, whereas GRI 417–419 cover three of the bottom four spots within PR and among all social GRI Standards. GRI 415 (Public Policy) ranks third at 7.28% (CV pf 0.21) and attempts to quantify the contribution Cadeler A/S makes towards the achievement of political goals per country and beneficiary operated in (Global Report-

[5] See tables 1 and 9 in the appendix.

[6] See tables 1 and 10 in the appendix.

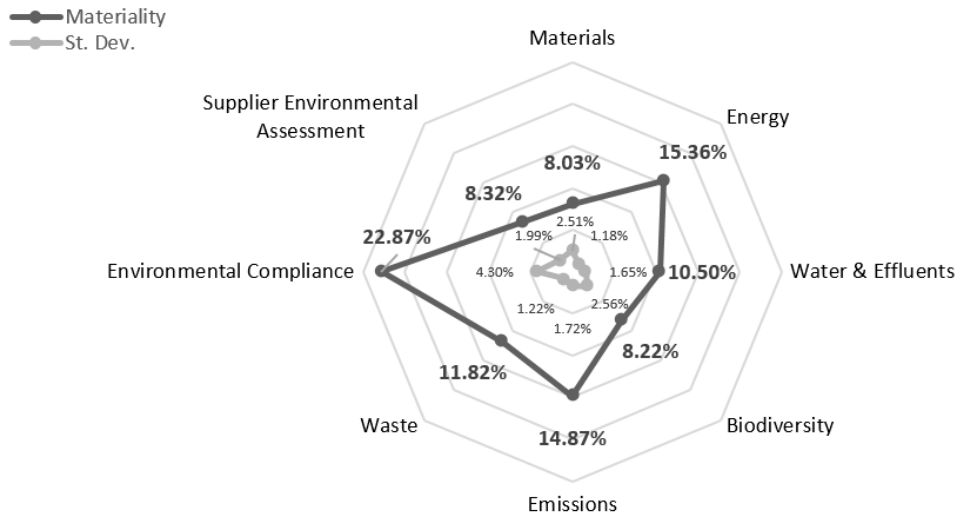


Figure 8: Materiality and Dispersion of Environmental GRI Topics (own work)

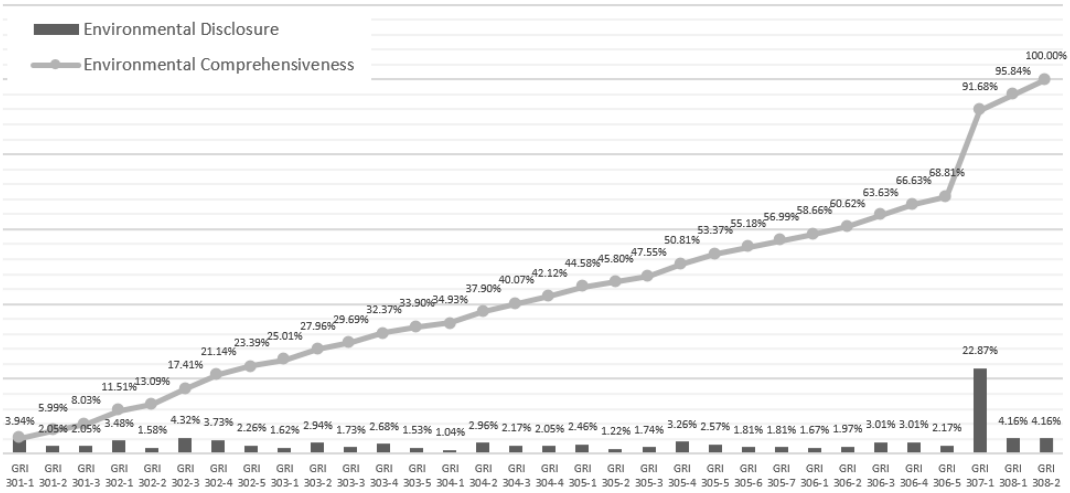


Figure 9: Comprehensiveness and Materiality of Environmental GRI Disclosures (own work)

ing Initiative & University of Stellenbosch Business School, 2020). Such contribution not only includes corruptions risks or lobbyism, but also party financing (ibid.).

GRI 415 ranks highest within the SOC subcategory and is closely followed by GRI 414 (Social Supplier Assessment) within the same subcategory, with that topic ranking fourth across all social GRI Topics at 7.21% and a below-average CV of 0.17. This topic deals with HRDD conducted on new suppliers that is directed to negative social impacts that can adversely influence activities in Cadeler's value chain or a large commissioning project. The fourth social subcategory, HR, was found to be the most important subcategory in the assessment at 31.32%. However, the category's two highest-ranking topics, Child Labor (GRI 408) and Forced or Compulsory Labor (GRI 409) take the global ranks six and eight at 6.47% and 5.88% and CVs of 0.53 and 0.54.

Regarding social disclosures, the results for GRI 403 show that 403-9 and 403-10 entail the highest local (global) materiality levels at 13.60% and 14.00% (1.18% and 1.15%) and deal with work-related injuries and illness. Its inherent management approach disclosures 403-1 to 403-7 are found to range lower and between 0.60% and 0.95% although the GRI Standards mandate their prioritization. Within GRI 416, the disclosure GRI 416-2 is weighted locally (globally) at 70.83% (5.40%), making it one of the highest-ranking social subtopics. It provides guidance on reporting incidents linked to a product's or service's noncompliance caused by shortcomings in the safety concept of operations with clients (see Global Reporting Initiative & University of Stellenbosch Business School, 2020).

Similar to GRI 307 or GRI 206, GRI 415 also covers one disclosure element, GRI 415-1, at a global weight of 7.28% that deals with contributions directed towards political causes through parties, beneficiaries, or governments that

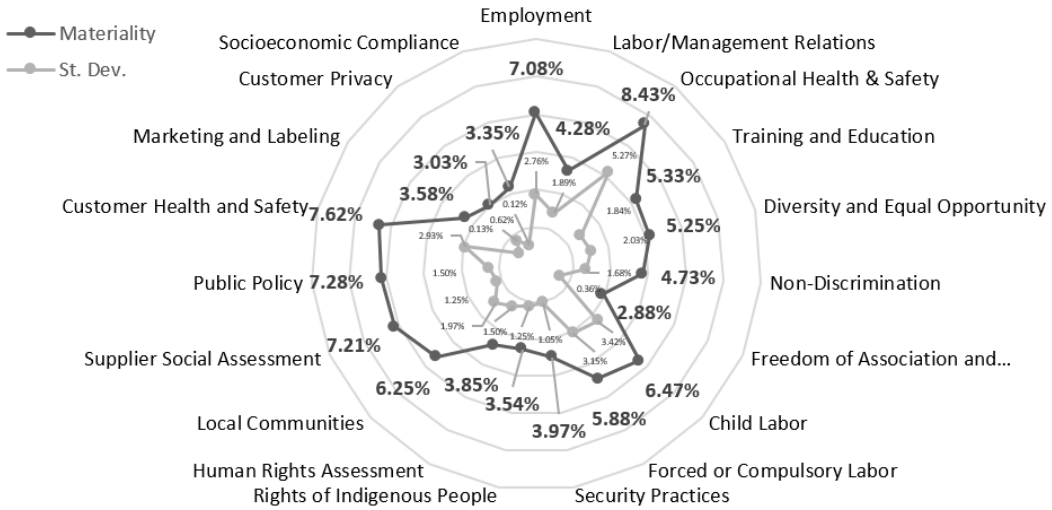


Figure 10: Materiality and Dispersion of Social GRI Topics (own work)

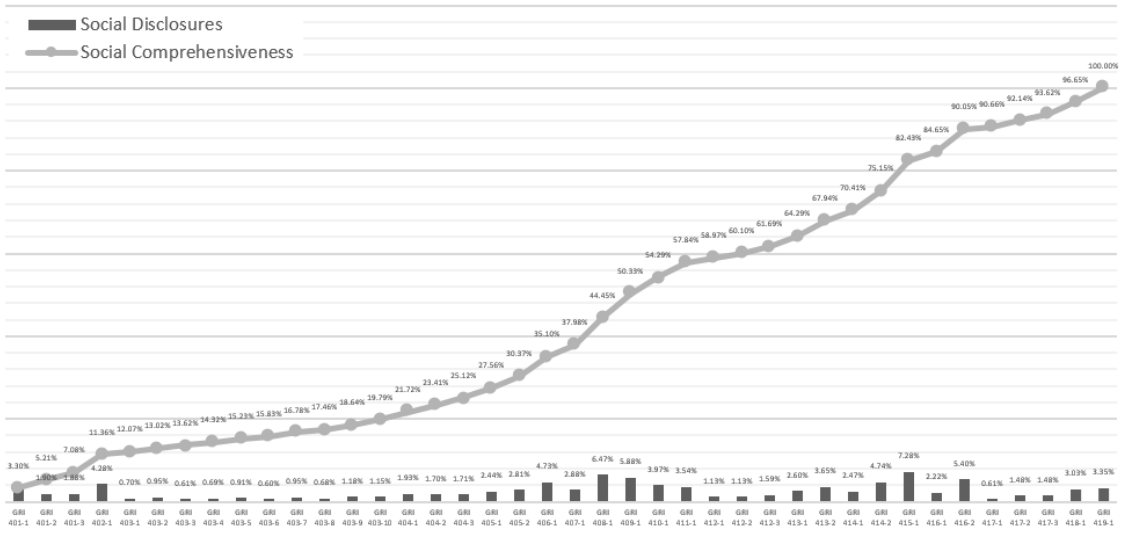


Figure 11: Comprehensiveness and Materiality of Social GRI Disclosures (own work)

can either be captured in the form of monetary payments or estimates, e.g. the value an offshore windfarm commissioning program or other services provided to a party's political agenda (Global Reporting Initiative & University of Stellenbosch Business School, 2020). Similar to GRI 416, the second disclosure in GRI 414 is prioritized at 4.74% globally and emphasizes occurrences of and actions taken against value chain activities that impede or adversely influence an organization or project's ability to create dual and shared value (Global Reporting Initiative & University of Stellenbosch Business School, 2020; Porter & Kramer, 2011).

Within Employment (GRI 401), main emphasis lies on new hires and turnover (GRI 401-1), which covers 46.64% of GRI 401 and 3.30% of all social disclosures. GRI 401-2 and 401-3 that deal with benefits provided and parental leave provided and taken are ranked almost identically at 1.90% and 1.88% globally. GRI 408 and GRI 409 entail only one disclosure item each that (1) requires disclosure on operations and suppliers at risk of child labor, incidents of such, and abolition measures (GRI 408-1), and (2) covers operations and suppliers at risk and/or showing incidents of forced and/or compulsory labor and mitigation measures taken. Both disclosures' weights are equal to their topics' (6.47% and 5.88%).

4.3. Stakeholders' Priorities

Along with time constraints in the data collection process, a major concern that arose in Cadeler's management is that direct inquiries regarding external parties' reporting preferences may lead to expectations that Cadeler A/S may not be able to live up to in a satisfactory way due to information asymmetries in expectations and strategic considerations. Therefore, external stakeholders' SR priorities towards Cadeler were inferred from manual content analysis

on stakeholders' sustainability reports and surveys of which the approach is outlined in section 3.2. Although the nature of this approach is seemingly not in line with the GRI's notion of stakeholder inclusiveness transcending the identification, prioritization, and verification of material content elements, the standard setting organization remains silent on a specific definition of such and thus gives room for interpreting stakeholder inclusiveness. Among, direct competitors, only one organization is listed in the GRI's database yet does not provide GRI-compliant SR. Therefore, the inference from reported material topics to GRI topics is a "most likely" translation from proprietary to framework jargon as no organization in the sample reports compliant to the GRI. However, offshore windfarm commissioners tend to have similar stakeholder groups, so their materiality assessments can be used to proxy expected stakeholder interests in Cadeler A/S.

The results displayed in Table 11 in the appendix show that among direct competitors, GRI 205, GRI 206, and GRI 207 recur most frequently, whereas the remaining topics are not found to be covered in the SRs. OEMs rank GRI 205 and GRI 206 highest and only one organization reports on procurement practices (GRI 204). As with direct competitors, the remaining economic topics were not found to be deemed material under sustainability concerns. Spearman's rho is low at 0.3482 and insignificant at a 95% confidence level. Windfarm developers, taken together, tend to prioritize GRI 201, GRI 205, and GRI 206 and, on the group level, exhibit a prioritization very similar to that of Cadeler A/S. The author finds a highly positive and significant correlation between the economic GRI priorities provided by Cadeler A/S and windfarm developers of 0.8125 and a p-value of 0.0264. On the audience level, GRI 205 and GRI 206 are slightly prioritized over GRI 201, and the author finds a strong positive yet weakly significant association.

Table 12 in the appendix displays the results from the analysis on environmental topics. Direct competitors tend to consider Emissions, Energy, and Environmental Compliance (GRI 305, GRI 302, and GRI 307) as most material and report on the remaining five environmental topics when taken as a group which are ranked similarly to Cadeler's. This is underlined by a highly positive and significant Spearman rank correlation coefficient of 0.8095 and a p-value of 0.0149. The opposite conclusion is drawn for OEMs, which tend to provide more balanced environmental reporting yet prioritize Emissions, "Waste" (GRI 306), and "Supplier Environmental Assessment" (GRI 308). Windfarm developers consider GRI 305, GRI 302, and GRI 308 material most frequently, and the topic ranking shows a weaky positive and insignificant association with Cadeler's priorities. On the audience level, the positive association is 12.5 percentage points weaker than that for direct competitors and is weakly significant (p-value of 0.0611). On a side note, all environmental GRI topics have been referred to in the sample of stakeholders' SR material.

The results on the analysis directed to social topics in Table 13 in the appendix show that Cadeler's direct competitors are most likely to prioritize Occupational Health and Safety, Non-Discrimination, and Diversity and Equal Opportunity with them recurring eleven, seven, and six times, respectively, with the first rank identical to that of Cadeler A/S. Social topics falling under the PR subcategory are not covered in the materiality analyses except for GRI 416, which was found in one competitor's materiality analysis. Spearman's rho is found low and insignificant. Among OEMs, GRI 403 is ranked highest and followed by the GRI Standards GRI 413, GRI 414, GRI 401, GRI 402, and GRI 405. Among PR topics, only GRI 416 is referred to once. Interestingly, Training and Education (GRI 404) is not referred to by any OEM's materiality assessment. Despite a medium magnitude in Spearman's rho, the association between the ranks is found to be significant. Windfarm developers tend to refer to GRI 403, GRI 404, GRI 405, GRI 401, and GRI 406 most frequently in this order and tend to refer to topics within SOC at a medium frequency and to topics within PR with the least, except for GRI 416. Spearman's rho is slightly less than in the OEM group with 0.4382 and weakly significant. Taking the stakeholder groups together, the qualitative ranking of top five topics is very similar to that of windfarm developers, and Spearman's rho of 0.3873 is found insignificant.

To summarize, two key observations are made. First, medium to high and significant correlations are found for only one group of stakeholders within the sample. For economic topics, it is windfarm developers whereas for environmental and social topics, the findings tend to not result from the method or sample when considering direct competitors and OEMs, respectively. Second, the two other stakeholder groups in the considered categories exhibit either weakly significant or mostly insignificant associations. Topic rankings inferred from the entire stakeholder sample tend to be either weakly significant or insignificant.

4.4. SDG Integration

This section looks at how Cadeler's reporting priorities across economic, environmental, and social topics could translate into SDG coverage along with their respective targets. As outlined in section 3.3, the translation is displayed on the topic level; however, the SDG targets and SDGs are sorted based on the GRI disclosures' local materiality. That is, the more material a GRI disclosure is locally, the earlier are the SDG target and SDG listed in their respective column; the analysis is based on the preceding internal materiality assessment and guidance provided by Global Reporting Initiative (2020b) and United Nations (2015b). For brevity and to show a potentially realistic example on how Cadeler can leverage the UN's comprehensive framework that contains 17 goals, 169 targets, and 231 unique indicators (United Nations, 2015b), this section assumes that Cadeler A/S decide to only report on material topics that exhibit a materiality score above the average weight assuming DM indifference. Therefore, the following sections will look at economic, environmental, and social topics with global materiality scores of more than 14.29%, 12.50%, and 5.27%, respectively.

4.4.1. Economic SDGs

In the Economic category, the topics with above average materiality are GRI 201, GRI 205, and GRI 206. For the former, the translation manual suggests the SDG targets 13.1 for the disclosure GRI 201-2 and the targets 8.1, 8.2, 9.4, and 9.5 as covered by disclosure 201-1. SDG target 13.1 calls for climate change adaptation efforts, whereas the SDG targets linked to GRI 201-1 cover economic growth, productivity gains, eco-efficient value creation, and GHG emissions relative to value creation, and advancing research functions, respectively (United Nations, 2015b). Therefore, GRI 201 is linked to the SDGs, 8, 9, and 13. Regarding GRI 205, all three disclosures are associated with SDG target 16.5, which aims to decimate the presence of bribery (ibid.). Similarly, GRI 206 is linked to SDG 16, with its equivalent target 16.3 promoting the importance of lawful behavior. This SDG target does not explicitly link to economic compliance but rather to injustice that legal systems impose on non-convicted individuals (ibid.) which makes the link between the GRI and the SDG target more symbolic than accurate in terms of content. Taken together, above-average material economic topics cover four unique SDGs and six unique SDG targets.[7]

4.4.2. Environmental SDGs

GRI 307 and its disclosure GRI 307-1 are linked to SDG target 16.3 and thus SDG 16. Conceptually, the problem with this association is the same as with GRI 206 as the SDG focuses on lawful behavior of legal institutions rather than organizations exhibiting environmental compliance. Based on an additional content analysis, SDG target 12.4 (lawful chemicals and waste management to limit their adverse impact) seems more appropriate as it points at material handling in line with such ruling. The association tends to be conceptual in the sense that it only alludes to the idea of compliance in an open fashion. GRI 302 and all five disclosures are associated with the SDG targets 7.3, 8.4, 12.2, and 13.1, which cover the relative contribution of renewable energy to the global energy supply, resource efficiency, eco-efficiency in material management and use through reduced footprints, and climate change adaptation efforts, respectively (ibid.). Regarding GRI, its most material disclosure GRI 305-4 (GHG intensity) is linked to the SDG targets 13.1 (climate change adaptation), 14.3 (fighting ocean acidification), and 15.2 (foster sustainable forest management) (ibid.). GRI 305-1, the second most material disclosure, is associated with the same ones and 3.9 (reducing air pollution) and 12.4. In total, the defined range of environmental topics covers seven SDGs and nine unique SDG targets.[8]

4.4.3. Social SDGs

Social SDGs are prioritized as follows. To start with, GRI 403 is associated with SDG 8 with its target 8.8, seven unique targets 3.3 – 3.9 along with the SDG targets 16.1, and 16.7. The former target is most present within GRI 403 with its emphasis on increasing workplace security and reducing adverse work-related impacts. This target is also referenced most frequently with above-average material social topics (eleven times) and within GRI 403 (nine times). The SDG targets 3.3–3.9 are referenced to mostly once and are linked to reducing mortality and increasing health coverage by mitigating adverse health and safety impacts; though work-related safety is not mentioned explicitly, the link to these targets is less explicit and more conceptual. A qualitatively similar finding is related to the associated SDG targets 16.1 and 16.7, though the latter implicitly points at compliant decision-making directed to population safety, of which the work environment can be seen as a subset (ibid.). GRI 416 takes a client perspective on work safety; conceptually, SDG target 8.8 can expand the compliance-driven translation related to 16.3, which stresses the role of legal enforcement. GRI 415 is linked to SDG target 16.3 because both alternatives point at the mitigation of bribery risk and potentially resulting negative societal impacts.

With GRI 414 emphasizing risk arising from socially unsustainable supplier practices, the targets 5.2, 8.8, and 16.1 geared to eliminating workplace safety, partly by eliminating violent practices, the association is deemed an accurate translation for this standard. GRI 401-1, the most material disclosure in GRI 401, is associated with the SDG targets 5.1, 8.5, 8.6, and 10.3, of which 8.5 and 8.6 exhibit the most explicit link to GRI 401. The targets 5.1 and 10.3 emphasize non-discrimination and equal opportunity (ibid.). The target 8.6 alludes to training and can thus be added to GRI 404, which shows strong overlap with GRI 401 and adds the SDG targets 4.3–4.5 and 8.2 that deal with increased access to education and productivity gains as such. GRI 408 and GRI 409 also exhibit substantial overlap as both are associated with SDG target 8.7 (banning forced and child labor) and the former on the target 16.2 (end child abuse). Finally, GRI 413 is linked to the SDG targets 1.2 and 2.3 which aim to reduce poverty by increasing work productivity across sectors and thus income, wealth, and SD in communities. Taken together, Cadeler's above-average material topics cover eight SDGs and 27 SDG targets, of which SDGs 8, 3, and 16 and the targets 8.8, 8.5, 16.1, and 5.1 are covered most frequently.[9]

4.5. Sensitivity Analysis

Robustness tests changing the base case assumptions test the results from section 4.1 for ranks reversal by simulating collective DM bias and exposure to uncertainty (Majumdar, Tiwari, Agarwal, & Prajapat, 2021; Velasquez & Hester, 2013). It mimics DM bias by manipulating a TFN's center of gravity towards its upper or lower bound by varying λ and/or a DM's uncertainty in preference setting by varying the vagueness of/overlap between linguistic options through α (Tsai et al., 2010; Tseng & Lin, 2008). One parameter is changed at a time for each DM and tested for consistency.

[7] See table 14 in the appendix.
[8] See table 15 in the appendix.
[9] See table 16 in the appendix.

Pre-tests for all 20 cases show that all DM's CRs remain below 10%. Therefore, the same number of PCMs is aggregated. The author uses the words "pessimism" and "optimism" with their variations synonymous to DMs' partiality to the lower or upper bound of their linguistic judgments' TFNs.

4.5.1. Economic Standards

Rank reversal is not observed when introducing DM bias to the base case. As expected, materiality scores are identical when $\alpha = 0$ as the conversion scale becomes crisp. Across all variations of fuzziness when keeping DM bias constant, the materiality scores converge slightly, meaning that the lower-ranking economic topics such as GRI 207, GRI 204, or GRI 202 gain relative importance at the expense of higher-ranking items GRI 201, 205 or 206. In the case of $\lambda = 0$, rank reversal is not observed when increasing uncertainty when compared to the base case in which $\alpha = \lambda = 0.5$. In the unbiased setting, GRI 203's rank increases from five to four whereas that of GRI 204 decreases accordingly when fuzziness is high or extreme with $\alpha \geq 0.8$ or rather when the overlap between linguistic judgments is highest and DM bias absent. In the setting with $\lambda = 1$ and α increasing, changes in ranks are observed on three topics when $\alpha \geq 0.4$. To illustrate, GRI 203 and GRI 207 trade the fourth and fifth rank $\alpha = 0.4$ as the latter's weight decreased from 13.03% to 12.69% and the former increasing above. At $\alpha = 0.8$, the rank of GRI 204 improves from six to five whereas GRI 207 loses one rank at a materiality score of 12.37%.

Holistically, the four highest-ranking economic topics lose relative importance with increasing fuzziness, with GRI 207 losing one spot in the unbiased setting and two spots in the optimistic case, whereas the alternatives that were ranked fifth or sixth in the base case gain materiality. GRI 202, the lowest-ranking items, gains most relative importance yet remains the lowest-ranking topic in all scenarios. High to extreme levels of fuzziness paired with small differences in medium-ranked materiality scores and DM optimism indicate sooner rank reversal and stronger convergence. For instance, GRI 201's materiality decreases from 18.69% to 17.63% when increasing fuzziness, whereas the decrease is less for when bias is excluded and even slightly negative when preferences are pessimistic ($\alpha = 0$). Importantly, the ranking of the three highest-ranking topics remains unchanged across all levels of fuzziness and bias. 13 of 20 rankings are equal to the reference and rank changes are only observed between the ranks four and six.[10]

Disclosure convergence within economic topics is observed for most cases except for those in which disclosures are ranked equally important or in which there is only one disclosure. The lowest-ranking disclosures within a topic gain relative importance at the expense of their highest-ranking alternatives. For instance, at $\alpha = 0.5$, the materiality score of GRI 201-1 decreases from 40.22% to 36.86% when increasing fuzziness from 0 to 1, whereas GRI 201-3 increases from 11.3% to 14.90%. Across all topics that contain heterogeneously ranked disclosures, rank reversal is found absent.[11]

4.5.2. Environmental Standards

The results on environmental topics show that their relative importance converges for every level of DM bias and increasing fuzziness. Likewise, the rankings of the three most material topics remain robust across all variations of DM bias and uncertainty and the ranking is robust to changes in DM bias at $\alpha = 0.5$. When $\alpha = 0$, the ranking remains robust until GRI 303 and 306 switch ranks when $\alpha \geq 0.8$. In the unbiased and optimistic setting, GRI 304 and GRI 308 exchange the ranks six and seven. In the unbiased setting, rank reversal also occurs only if $\alpha \geq 0.8$ yet two different topics, GRI 303 and GRI 308 exchange the sixth and seventh rank. The same observation is made at the same minimum level of fuzziness and $\lambda = 1$. In total, rankings from 14 of 20 scenarios replicate the base case.[12]

Disclosures within GRI 301, GRI 307, and GRI 308 are found robust to changes in DM bias and uncertainty across all scenarios. Priorities within GRI 304 are found to be robust in all scenarios except for when $\lambda = \alpha = 1$ in which GRI 304-3 and 304-4 exchange the third and fourth rank. Priorities within GRI 302 only exhibit rank reversal when $\lambda = 0$ and $\alpha = 1$, where, interestingly, GRI 302-5 and GRI 302-4 switch the first and second rank. Regarding GRI 303, 13 scenarios are robust to variations with ranks changing most frequently when $\alpha \geq 0.8$ across all types of bias. In the case of GRI 305, 14 scenarios are consistent with the base case, with most rank reversals taking place at medium to low ranks at $\alpha = 1$ across all bias levels and with $\alpha \geq 0.6$ when $\lambda = 1$. Across GRI 306, only medium to low-ranked elements change priority when $\lambda = 0$ and $\alpha \geq 0.8$. In total, disclosure rankings are robust when changing fuzziness in the TFNs and when $\alpha < 0.5$ across all bias levels.[13]

4.5.3. Social Standards

Table 21 in the appendix shows the results on the subcategory level and that their ranks remain stable across all bias-fuzziness combinations. The results on social topics in Table 22 in the appendix show that ranks do not only change among medium- and lower-ranked topics in the prioritizations, but also among high-ranking alternatives though only when fuzziness is extreme. Specifically, GRI 403 and GRI 416 are ranked first and second in all scenarios except when $\lambda = 0$ and $\alpha = 1$. The only topics exhibiting robust ranks across all scenarios are GRI 405, GRI 407, GRI 409, and GRI 418. GRI 402 to GRI 404 show variation when introduced to bias and extreme fuzziness at $\alpha = 1$. Excluding the base case, the topic ranking is perfectly robust in the unbiased setting when $\alpha = 0.6$ and in the optimistic setting when $0.5 \leq \alpha \leq 0.6$. In addition, the results show that the frequency and extent of

[10] See table 17 in the appendix.

[11] See table 18 in the appendix.

[12] See table 19 in the appendix.

[13] See table 20 in the appendix.

deviations is mainly contingent on α rather than λ. To illustrate, whereas five topics deviate from the base case when assuming $\lambda = 0$, the number remains unchanged when making TFNs crisp and, on average, doubles at high and extreme fuzziness levels. In the unbiased and optimistic cases, manipulating default fuzziness tends to lead to more ranking variability.

All disclosure preferences from GRI 412 upwards are robust across all scenarios, whereas all disclosures below show rank variability. Disclosures within GRI 405 and GRI 401 only exhibit priority changes at $\alpha = 1$ and $\alpha \geq 0.8$, respectively, across all types of bias. GRI 404-1 is only found to change its top rank when $\lambda = \alpha = 1$ whereas its lower-ranked alternatives change ranks with decreasing fuzziness in the unbiased and pessimistic scenarios and extremely low or high levels of fuzziness in the optimistic case. Lastly, findings on GRI 403 show that its most material disclosures GRI 403-9 and GRI 403-10 are robust across all combinations, with the latter sharing the same rank with the former when $\lambda = \alpha = 1$. GRI 403-1's rank is robust to zero bias and optimism and $\alpha \leq 0.2$ across all λ. The remaining disclosures show mixed findings yet are more susceptible to rank changes with increasing fuzziness and DM bias. Combining optimism and low deviations from medium fuzziness results in the least rank variation relative to the base case.[14]

Table 24 in the appendix provides a summary of the preceding analysis on topic and disclosure rankings and shows the scenarios', items' (topics and disclosures), and rank's accuracy in replicating the base case. Accuracy is defined as the percentage of cases that are robust across all topics, disclosures, and scenarios, respectively. When a social topic includes more than one disclosure, their total is reflected in the number of total disclosures across all bias-fuzziness combinations. The results show that most scenarios accurately represent economic topics and disclosures, whereas most economic topics and disclosures are robust across all bias-fuzziness combinations. In addition, social subcategories, topics, and disclosures when ranked on the subcategory level exhibit high accuracy, unlike the aggregate of all described previously. Finally, the ranks of environmental topics and economic disclosures are most persistent globally.

5. Discussion & Conclusion

The objectives of this research were to shed light on how Cadeler A/S could plan their future SR activity by deploying a quantitative materiality assessment process as a resource planning tool that provides the organization with a structured approach to prioritize their allocation of financial, cognitive, and temporal resources towards SR content elements and related data collection efforts and long-term progress towards the UN's SDGs. The research questions addresses in this study were geared towards (1) how Cadeler A/S could, with a structured and scientific methodology, prioritize SR topics at limited organizational information-processing capacity that makes the resulting SR materiality-driven, compliant with established SR regulation, endorsed standards, and can substantiate contribution towards the UN's SDGs, and (2) how Cadeler A/S could go forward to integrate such materiality-driven SR in the organizations to make sustainability data processing more responsive, inclusive, and ubiquitous to reconcile operative efficiency with an ambitious growth trajectory observed in the offshore windfarm construction market (European Commission, 2020a). One of the main contributions of this thesis is that it empirically determines materiality scores for an SME rooted in the offshore wind sector and its stakeholder audience and provides a suggestive agenda for how Cadeler A/S can enhance their future SR in an emerging industry to increase their chances to secure longer-term success through informative disclosure activity.

This study applied Group-based Fuzzy AHP based on Calabrese et al. (2016) and is paired with a manual content analysis screening SRs of a sample of 19 stakeholders linked to Cadeler A/S that are segmented into groups of direct competitors, OEMs, and windfarm developers. The analysis meets a dual purpose as a tool for (1) both prioritization and resource planning centered around the preparation of a GRI-compliant report and (2) setting a comparative case for reporting preferences local to Cadeler A/S and across the larger-scale organizations that are members of the offshore wind supply and value chain as organizations winning contracts, providing energy-converting assets, and delivering the construction service to the second. This section outlines the key findings from the preceding analyses and looks at the implications for Cadeler regarding the research questions posed on this report. On a general note, what Cadeler A/S may take as a given or required by law may be considered highly material by external stakeholders who are less familiar with intra-industry focal points.

5.1. Commissioning SR

The results from the economic section suggest that Cadeler A/S should run a four-pronged approach centering around Economic Performance including economic value added and climate-related financial and value risks, integrity in business conduct through GRI 205 and GRI 206 by shedding light of bribery risk and trials based on related allegations, and emphasizing tax matters such as their approach, governance, and stakeholder management related to such. Priorities within GRI 207 show that it is consistent with the GRI's prioritization of the topic's management approach disclosures. The results further show that these topics' respective CVs exhibit the highest dispersion among DMs' individual judgments. Interestingly, individual submissions were not necessarily in line with the author's expectation that managers prefer what they are most familiar with, such as the submission from one DM specialist in procurement (Schläpfer & Fischoff, 2010). Insights like these indicate that respondents may have considered the broader scope of

[14] See table 23 in the appendix.

organizational reporting priorities rather than local phenomena on the department level. This partly rejects the implicit hypothesis that returned questionnaires entail department bias (ibid.).

Considerable dispersion among high-ranking topics and notably within GRI 207 can be explained by one DM ranking it highest, whereas two other DMs ranked it in the lower half of the topic ranking at half of the materiality score. GRI 205 and GRI 206 are ranked highest by two DMs whereas one DM ranked them as the second-lowest priorities. Besides possible prioritizations for SR planning and construction, the results across all categories confirm T. C. T. Chen (2020) in that group-based decision analysis with FAHP is susceptible to lacking perfect consensus and that aggregating results may over-simplify internal preferences, which may also be driven by content elements' perceived reportability rather than their strategic importance to Cadeler A/S, or managers' personality traits or affiliation, since the sample consists of DMs reporting to the CEO. Although a participative approach intends to mitigate the risk for possible impression management, the survey cannot perfectly mitigate it despite capturing verbal preference cues. Untabulated results showing DMs' individual preferences indicate that a topic's CV increases when most topic-level materiality scores are at least 50 percent higher in magnitude than the minority's vote. The correlation between the materiality scores and the CVs at 0.2832 points at a weakly positive and statistically insignificant relationship across all topics. Dedicated sensitivity analyses on topics and disclosures show that the preferences, on an aggregate level, are highly robust to changes in collective DM bias and exposure to uncertainty and informational value discounts (Dubois & Prade, 1988).

High CVs in economic priorities can render task planning and delegation more complex by pointing at control hotspots and greater need to coordinate DMs with alike preferences to render collaboration more resource- and cost-efficient (George et al., 2016; Herremans & Nazari, 2016). To illustrate, DMs with higher preference similarity to each other and top management's interests may exhibit a higher willingness to collect and report related data for internal information-processing and require a stronger involvement in organizational action centered around their inclination. Such actions tends to fall in line with Herremans and Nazari (2016) who advocate delegating specialist tasks to the most committed experts with the most fitting perspective. This would require screening more executives' preferences and higher procedural legitimacy driven by a promoting member or group (see Kiesnere & Baumgartner, 2019). Accomplishing this outcome, however tends to hinge on a culture responsive to organizational change and top management support towards creating stronger linkages between HSEQ, management functions, and organizational units across all hierarchy levels (ibid.; Adams & Frost, 2008; George et al., 2016).

Broadly speaking, most of Cadeler's stakeholders tend to put little emphasis on reporting on economic topics which confirms assertions in previous studies on a propensity towards social and environmental reporting (Hubbard, 2009; Saenz, 2019). It is likely that related disclosure can impact the perceived relative importance of social and environmental reporting (ibid.). The reliability of this data is also impacted by that the priorities are derived from disclosure material that may not directly address Cadeler A/S or can be translated to a GRI equivalent. For instance, reporting GHG emissions based on the MARPOL conventions is a common requirement to Cadeler A/S stated in dedicated sustainability surveys, though SRs emphasize the topic rather than the method or policy. A higher-level description tends to keep disclosure material more concise at the potential expense of explanatory power; over time, more detail will have to be disclosed either on demand or as a supplement provided at the expense of potentially less convenient access and conciseness. Put differently, an organization adopting SR is likely to have to trade off conciseness and comprehensiveness in a longer-term setting assuming scale increases (Zhou, 2011).

The observed lack of economic coverage can come from at least two possible sources. First, economic topics may be important for an organization's sustainability yet were omitted in the report, which would be consistent with (Beske et al., 2020) pointing at impression management to signal stronger commitment to nonfinancial topics besides economic considerations. Given that neither direct competitor complies with the GRI Standards, this could explain why the author finds little to no information on how these organizations constructed their materiality matrices and defined the topic lists. This finding also holds for two out of three OEMs and five out of ten windfarm developers, and confirms Machado et al. (2021) in that GRI-compliant reporting on the materiality assessment methodology is mostly high-level in the absence of regulation enforcing such practice.

Given the significant reporting gap in that no direct competitor in the sample adheres to the GRI framework, this can represent a strategic opportunity for Cadeler A/S to adopt economic GRI standards signal commitment to informative SR towards OEMs and windfarm developers, of which all organizations in the sample engage in SR in line with the GRI Standards and also investor mindset as the adoption of such framework is shown to be positively linked to SR quality and sustainability performance, as found in Michelon, Pilonato, and Ricceri (2015), stock price performance gains (Guidry & Patten, 2010) and theoretically shape industry or market practice in SR in the presence of Cadeler's leading position in the offshore windfarm commissioning market on the EU level (see Moseñe et al., 2013). This strong disparity in reporting efforts and significance scores in preference alignment indicates that reported congruity with a stakeholder audience's preferences tends to assume away divergence among sub-groups' and/or individual reporting preferences through potential oversimplification, thereby confirming Puroila and Mäkelä (2019). Still, OEMs tend to emphasize non-economic topics and lawful business conduct, whereas windfarm developers show a relatively stronger propensity to report on economic performance. It should be noted, however, that the higher topic count can also be attributed to the sam-

ple size and potentially more comprehensive materiality matrices, though they were found similarly exhaustive among stakeholder groups with direct competitors, OEMs, and windfarm developers covering 17, 16 and 20 GRI Topics on average, respectively.

Economic SDGs are primarily driven towards integrity and access and commitment to lawful procedures and climate risk arising from Cadeler's underlying economic activity geared towards economic growth and value added for the organization, its employees, economic sustainability environment (i.e. industry and community) and innovation practices towards building eco-efficiency over time (Orsato, 2006). Therefore, Cadeler A/S is recommended to focus economic reporting around the SDGs (targets) 13 (13.1), 8 (8.1 and 8.2), 9 (9.1, 9.4, and 9.5), and 16 (16.3 and 16.5), adding the three latter items to Cadeler's SR portfolio (Cadeler A/S, 2021). Notably, the link to SDG 16 from the GRI Standards is rather conceptual as the SDG does not explicitly address compliance to economic regulation but rather access to justice. Reporting on such could be deemed optional since industry compliance can be regarded as a license to operate and not a competitive edge (Deegan, 2002; Global Reporting Initiative, 2020b). Therefore, organizations should not slavishly follow GRI recommendations but also review how well the SDG description fits the GRI element to avoid seemingly aligned reporting for the sake of signaling beyond-compliance behavior (Orsato, 2006; Power, 2009). In this case, however, the author deems the translation accurate and potentially worth reporting, though this decision should factor in a larger extent of managerial judgment.

The results on environmental standards show that DM consensus in Cadeler A/S considers Environmental Compliance, Energy, Emissions, and the newly introduced "Waste" standard most material when adopting GRI-compliant SR. This confirms Talbot and Boiral (2018) in that GHG emissions constitute a highly material topic due to its disclosure to a stakeholder audience. Contrary to the findings in the economic section, the most material topics exhibit higher consensus among DMs as indicated by their respective CVs, indicating that building common ground on constructing an agenda for GRI-compliant reporting may consume less of the resources that can be dedicated to data collection, building metrics, reporting, and stakeholder engagement (Baumann-Pauly et al., 2013; Buonocore et al., 2019; Whitehead, 2017).

Considering the materiality assessment on the stakeholder groups, direct competitors, taken together, tend to emphasize Emissions, Energy, and Environmental Compliance most frequently. The finding that compliance recurs less could be explained by either the sample (size) or rather by the mechanism in Deegan (2002) stating that compliance is an organization's license to operate which could be interpreted as a basic requirement that may not provide incremental informational benefit to informed readers. Follow-up interviews with Cadeler's sustainability management point at stringent and regular audits ensuring alignment to such policies as MARPOL or ISO requirements for certification and Green Club membership (Orsato, 2006). Since compliance is still alluded to rather frequently, one can infer that emphasizing compliance signals lawful business conduct to recipients who are less savvy on the importance of compliance in offshore wind or can be interpreted as an act of self-commitment for gaining legitimacy (see Boiral & Heras-Saizarbitoria, 2017).

The increased emphasis on emissions and energy tends to be observed across industries yet tends to be of strategic importance to offshore windfarm constructors in that WIV operation is an inherently transitional activity as it lacks a low-carbon alternative due to technological and cost constraints (EU Technical Expert Group on Sustainable Finance, 2019). Notably, prioritizing GHG intensity within GRI 305 can be linked to the organization's aspiration to build trust with committed investors and improve market valuation and eco-efficiency (Serafeim, Park, Freiberg, & Zochowski, 2020). Consequently, WIV operators can reap economic benefits from innovation in implementing eco-efficiency solutions and improve their sustainability performance and chances of living up to sustainability criteria required from windfarm developers contracting WIV operators which tend to span across membership to the UNGC or SBT or reporting emissions data to the CDP. These commitments, however, are only observed among OEMs and windfarm developers which can be explained by an increased administrative burden overcome with sufficient slack (Rasche et al., 2020).

In the case of windfarm developers, some reports' materiality sections were cut from one year to another without replacement. One can thus argue that using SRs from different years may be less timely input though more comprehensive and explicit on which elements a reporting organization tends to prioritize. The reasons for such change can be both explained by strategic omissions and thus drawing reader attention to different elements (Talbot & Boiral, 2018) as a measure to potentially oversimplify a materiality analysis (Puroila & Mäkelä, 2018), or for drawing attention to more positive content source (Beske et al., 2020). In any case, the materiality screening can become more derivative and prone to misinterpretation and comparability challenges within and across industries (La Torre et al., 2018; Torelli et al., 2020). Applying the findings to the sustainable business model taxonomy in Bocken et al. (2014) and Cadeler's SD report shows that Cadeler's business model tends to be primarily technological due to its strong emphasis on ensuring compliance and going beyond to increase resource/GHG efficiencies and circularity of operative and strategic input goods such as alternative fuels, MGO, or jacking grease (Cadeler A/S, 2021).

Concerning the SDGs, Cadeler A/S should be mindful of the type of compliance alluded to in the SDGs' description as environmental compliance is not covered by SDG 16, but rather dealt with by SDG 12 and its target 12.4. Further results indicate that Cadeler A/S should pay special attention to the SDGs 7, 8, 13, 14, and 15 when reporting against emissions, energy usage, and pollution. These findings are largely consistent with Cadeler's 2020 CSR report, which indicates less challenges in setting priorities than in measuring data or constructing sustainability control systems, co-evolving them

with traditional management controls, and driving organizational routines towards stronger SR integration and related performance evaluation (George et al., 2016; Gond et al., 2012; Kiesnere & Baumgartner, 2019).

The results on social standards assessment show that in the aggregate case, Cadeler A/S should focus most on Occupational Health and Safety, Customer Health and Safety Public Policy, Supplier Social Assessment, and Employment, respectively. Ranking GRI 403 first confirms Eccles et al. (2014) in that indicators on safety measures, incidents, and disclosures on safety management systems tend to be disclosed frequently across industries, including offshore wind. Whereas the focus on the former topic is consistent across all stakeholder groups, Cadeler A/S ranks GRI 416 second, whereas this topic is ranked in the lower half or third of social topics across the stakeholder groups. This finding has three implications. First, stakeholder organizations may consider the safety of a construction, manufacturing, or development service less important to report because safety procedures face extensive regulation in offshore operations or may associate such safety measures with GRI 403 applied to client visits on project sites. The underlying reason for paying little importance to such and the organization's assumptions stated in reports is sketchy and requires further inquiry. Second, it is likely that the subcategory layer in this study's approach drives this topic's relative importance as it is only compared in relation to its subcategory-level peers.

Third, differences in sample sizes and the comprehensiveness of each organization's sustainability report could drive the findings as the external screening method tends to reward more comprehensive and distinguished reporting. OEMs and windfarm developers tend to prioritize GRI 401 similarly to Cadeler A/S, whereas direct competitors tend to rank it lower. Similar observations of alignment in regards to GRI 415, GRI 414, and GRI 413 can indicate that social SR preferences relating to the more material topics are more aligned with suppliers of projects and related input material and potentially (1) a stronger commitment to signal compatibility between Cadeler and the individual stakeholder groups and (2) improve managing sustainability and reputational risk with the former's self-commitment to partners' standards (Anderson, Anderson, & Able, 2009; Green, 2015).

Using the aggregation layer to mimic modularity can, on the one hand, reduce decision-making effort by reducing the number required pairwise comparisons yet tends to introduce considerable variability to the model. Taking the subcategories one by one, LPDW should focus on GRI 403 and GRI 401, where HR should focus on GRI 408 and GRI 409. Within SOC, emphasis should lie on GRI 415 and GRI 414, whereas within PR, the main priority is GRI 416. The insights from the sensitivity analysis considering both a subcategory and an aggregate perspective indicate that the latter leads to a significantly stronger discount in the value of information used for decision-making which is reflected in low robustness and accuracy linked to more frequent changes in ranks not only across items with a medium priority but also with a high one (Dubois & Prade, 1988; İbrahim Özkan & Türkşen, 2014). The following example considers human rights coverage because anti-corruption, bribery, environmental, and employment topics requiring coverage according to the NFRD have been considered highly material across economic, environmental, and social categories (European Commission, 2014).

A rather surprising finding is that the aggregate materiality approach ranks mandatory human rights coverage in the middle of the topic ranking, whereas prioritization tends to look very different within subcategories. Cadeler's preferences on HR topics rank GRI 408, GRI 409, and GRI 406 highest; divergence in the aggregate ranking and that of their stakeholders is therefore likely to be driven by applying an intermediate subcategory aggregation layer that can either (1) state the relative importance of material items in alignment with Cadeler's strategic objectives, (2) overstate less material topics by applying the same subcategory weighting to them or promote the relative importance of subcategories with few items that are weakly prioritized over another, and (3) understate topics' relative importance due to a low subcategory weight or a larger number of items that require more material items to be assigned more relative importance. Put differently, the low robustness of social topics and disclosures on the aggregate level that mimics modularity by multiplying the subcategories' with their topics' local weights tends to stem from this aggregation step as it imposes a subcategory weight on every topic and disclosure that is part of it. This includes topics that are part of the same subcategory yet may be less important for Cadeler A/S to report than content elements from another subcategory. Consequentially, aggregate social priorities and robustness data should be interpreted and treated with caution prior to decision-making.

These pitfalls tend to limit the method's ability to compare alternatives across subcategories in terms of magnitude, though provide sensible guidance on how Cadeler A/s could prioritize content elements should they decide to cover all topics (disclosures) within a subcategory (topic). Practically, the topics ranked highest within "Human Rights", which constitutes a mandatory reporting element in the NFRD and Section 99a in Danish Accounting Law, are GRI 406, GRI 407, GRI 412, and GRI 408 for direct competitors, whereas the materiality sections of OEMs' reports show little coverage or preference for either topic (Authority, 2015; European Commission, 2014). Windfarm developers, on the other hand, also prioritize GRI 406, GRI 412, GRI 408, and GRI 409 among the highly material issues.

Depending on the policies applied for compliance to GRI 412, there is a chance that GRI 408 and GRI 409 are linked to this disclosure, meaning an adopting organization would have to consider these topics jointly. Such a process tends to be favored by delimiting the analytical scope to either "Human Rights" topics or by applying an additional pre-screening to the scope of GRI topics. The findings, however, could also be driven by reporting organizations' and their stakeholder audience's implicit assumption that reporting on incidences of child labor or compulsory labor may be strictly regulated and may not represent a major reporting concern resulting in less emphasis throughout the SR mandate, similarly to GRI

412. On the other hand, one can argue that even though the GRI Standards have a modular structure, some topics within subcategories may still be regarded as related which may be a remnant of the superseded GRI G4. For Cadeler, the procedure on prioritizing social SR can take two forms which are analogous to economic and environmental topics: They can either capitalize on global topics scores and consider above-average material items largely aligned with OEMs' and windfarm developers' priorities, or they could drill down the social hierarchy and select subcategory, topics, and disclosure priorities from their upper layers.

When looking at GRI compliance that may result in a "core" badge, the results for GRI 403 show that the two last disclosures on work-related incidents and illness are seen as most important. The disclosure ranking was conducted without priming DMs which disclosure would be required to report according to GRI to factor out desired compliance to that framework as a result driver. The findings are largely consistent with the organization's and industry's strong emphasis on Occupational Health and Safety represents an integral part of offshore windfarm commissioning and ensures that construction projects are completed on time and under highest standards (Cadeler A/S, 2021). Special emphasis lies on such information as procedures deployed to ensure work safety, programs intended to move beyond compliance, or the number of work incidents following noncompliance to HSEQ policies. The findings are consistent with the results obtained by Eccles et al. (2014) who find a positive link between organizational sustainability and reporting skill mappings and development strategies, the number of nearly missed and fatal accidents and the driving role of compliance to adopt such metrics, since HSEQ processes are frequently audited and certified to ensure operational safety (Cadeler A/S, 2021).

5.2. Handling Uncertainty and Stakeholders

In the sensitivity analyses, across all subcategories, topics, and disclosures, materiality scores tend to converge with increasing decision uncertainty across different forms of DM bias. Global preferences are mostly over 60% robust to changes in collective DM bias and TFNs' fuzziness. Convergence can be explained with the concept in Dubois and Prade (1988) that uncertainty, fuzziness, or vagueness introduced to a decision-making process tends to discount the value of the information used for setting a preference to compensate for undesired consequences from deciding on an alternative. This mechanism would also explain observed changes in ranks among alternatives that are much alike in terms of their materiality score. A higher-ranked alternative loses value (materiality) with increasing uncertainty which is used to compensate for relative importance that may have mistakenly been cut from a lower-ranked alternative. With similar rankings, a change in ranks thus becomes more likely and may appear at a lower level of uncertainty in the presence of DM bias, which, to an extent, tends to decrease the value of information. This would imply that an unbiased perspective in a fuzzy setting tends to entail more decision-making uncertainty than a pessimistic setting, which would assign higher value to the informational situation.

What remains unclear is which case is the more truthful scenario. Although crisp figures are a desirable means of simulating a higher confidence level in decision-making (Dubois & Prade, 1988), there is a possibility that its certainty could be constructed and may not accurately proxy the degree of (un-)certainty on the organizational level. Importantly, the observed trends can also be driven by assuming that all DMs aggregate share the same bias and attribute the same value to the information they use for preference setting (Calabrese et al., 2016). When introducing heterogeneity to the pool of DMs, it is likely that, on average, uniform extremely high or low decision-making uncertainty may not be representative of the DMs' collective informational perspective. Varied stances may result in a more balanced picture that could be more resemblant to the base case with different biases or uncertainty levels (partially) cancelling out. Present the low consistency ratios in DMs' preferences and that most respondents rarely used the more extreme ends in the linguistic preference scale, Fuzzy AHP tends to be the more appropriate means to capture different interpretations of verbal cues which classical AHP would assume away (Chan, Sun, & Chung, 2019). The method's ex-post suitability, in turn, could have been driven by cultural factors (Scholtens & Sievänen, 2013). Fuzziness (vagueness) in responses, on average, tends to positively drive rank changes in medium-ranked alternatives more than changes in DM bias alone. Varying DM bias and fuzziness simultaneously tends to accelerate rank reversals among the same alternative and partly to high-ranking priorities.

5.3. Implications for Management

The study suggests several measures for Cadeler's management to drive the organization from a transitional state between compliance-driven and peripheral sustainability integration beyond the latter and a synergistic ecosystem between management and sustainability controls and measurement used for incentive management and enhanced through employees' feedback. Given that the materiality assessment is based on topics that can be voluntarily adopted, the baseline GRI Standards 101, 102, notably 102-46 and 102-47, and GRI 103 along with detailed reporting on Cadeler's sustainable business model should be covered to build an informative and transparent foundation for materiality-driven and GRI-compliant SR (Bocken et al., 2014; La Torre et al., 2020; Machado et al., 2021). Importantly, the scope of material topics is not limited to GRI elements and can (and should) be extended to industry-specific issues to render proprietary SR more comparable. Doing so is recommended by setting a benchmark priority list based on Cadeler's sustainability strategy. Top managers will have to be engaged more extensively to treat control hotspots arisen from preference screening towards stakeholder centricity (Adams & Whelan, 2009).

Increasing technical knowledge and sensitivity among top managers to manage control hotspots to optimize sustainability information processing across hierarchical levels can

increase HSEQ's influence on SCS integration (Adams & McNicholas, 2007; George et al., 2016) and leverage specialist knowledge and learning economies with aligned interest to capitalize on higher willingness to engage on task specifics rather than over-aggregating preferences concealing them (Herremans & Nazari, 2016). Aligning control design towards truthful SR should be incentivized more intensely in reward systems (Battaglia et al., 2016).

Performance measurement systems are recommended to be equipped with GRI metrics to promote efficient decision-making towards legitimacy-building SR, alignment with salient stakeholders' priorities, and goal congruence fostering lower-cost decision-making (Hubbard, 2009) and complementarily to the present focus on HSEQ, Human Resources, and a broad scope of environmental metrics to be substantiated with more technical detail (Cadeler A/S, 2021). Impact-weighted accounts could potentially improve the explanatory power of processes and reported sustainability information (Serafeim et al., 2020). Importantly, expansion in fleet, project scope and requirements in information-processing should be accompanied by increases in related capacity, headcount, and communication channels and to balance the scale of the organization at the point at which coordination costs become substantial and growth detrimental to performance due to failure to incorporate learnings from expansion paths (Okhuysen & Bechky, 2009; Vermeulen & Barkema, 2002).

External stakeholder engagement will likely increase in relevance and result in more comprehensive SR. Still, the author recommends starting with the most material content elements and respond to stakeholder feedback by providing informational supplements or more comprehensive report. Assuming a longer-term horizon and a growing stakeholder audience, it seems unlikely that increasing SR comprehensiveness can be avoided without evoking thoughts of impression management and involving stakeholders (Beske et al., 2020; Jespersen & Olmsted, 2019). Importantly, the suggested priorities are dynamic over time in response to underlying operations and changing stakeholder groups. To limit cognitive effort, it is recommended to assess them for their salience (Mitchell et al., 1997). On the other hand, doing so keeps SR more subjective and potentially oversimplifies complex stakeholder relations; given the shortcomings of over-aggregated materiality matrices, it is recommended to balance segmenting stakeholder groups with preserving competitive edges (Machado et al., 2021; Puroila & Mäkelä, 2019). Changes in methodology, replacements and omissions, and reference points should be quantified through adjustments or avoided, respectively, to preserve SR's comparability, traceability, and transferability to a different state of Cadeler A/S (Puroila & Mäkelä, 2019; Talbot & Boiral, 2018).

To stress beyond-compliance behavior, it may be sensible to look beyond the GRI and NFRD and consider reporting against the EU Taxonomy's TSCs. As a recent development, legal guidance will become mandatory in late 2022 yet is presently unspecific on offshore wind. However, applying the latest draft material on their own and competitors' NACE Codes can provide valuable insights into data requirements and filing requests for review in case there is a misfit between the requirements and their assigned NACE code. So far, only a handful of windfarm developers with a longer history in and more resources for SR have adopted the policy that may become a selection requirement for future offshore windfarm construction projects. Further requirements could be include a GRI reporting badge or UNGC membership for signaling further commitment at the cost of an increased administrative burden (Rasche et al., 2020).

Depending on the market location, Cadeler A/S may also consider adopting a more investor-oriented and ready-to-use framework such as SASB to engage overseas stakeholders and in the presence of offshore wind gaining significant uptake in the US. Entering assurance engagements despite its cost and potentially emerging risk factors in the assessment can reduce long-term litigation risk and render information collection and processing more tailored to the organization's underlying activities and more efficient (Porter & Kramer, 2011). Finally, though compliance is integral to functioning operations and legitimacy building (Deegan, 2002), Cadeler should refrain from overemphasizing it to avoid appearing as risk-averse and compliance-driven, and instead incorporate climate risk scenario analyses into their reporting under the TCFD's recommendations, which pose further challenges to determining material climate risks and verification by external assurers (O'Dwyer & Unerman, 2020; Power, 2009). A structured materiality assessment should advance to an organizational routine carried out periodically to update Cadeler's sustainability strategy and refine responsiveness to their stakeholders' needs (Beske et al., 2020).

5.4. Limitations

The following limitations pertain to the study. First, the structured materiality approach only considers the scope of the GRI Standards and is aimed at finding out which minimum reporting requirements to prioritize if Cadeler adopts this framework. There is a risk that this approach omits (may include) potentially important industry- or investor-level topics that should have been added (removed) to (from) the framework due to their relevance (temporal inapplicability). In addition, aggregating stakeholders into groups can be silent on more specific requirements that may go beyond the GRI. Undertaking these changes would add substantial subjectivity to the analysis and may provide a potentially inaccurate objectification or construction of perception, which is why a foundational structure was imposed for the analysis. This includes the introduction of managers' relative importance, which was assumed equal yet can also have driven the results to a considerable extent. Similarly, there is a possibility that when setting their preferences, managers took operative and strategic perspectives that are not captured by the survey. Subjectivity remains inherent to the materiality assessment and is unlikely to be ruled out when founded on managerial judgment (Beske et al., 2020). The analysis further assumes similar DM bias and value attributed to information that may both vary considerably. Importantly,

Cadeler's present reporting and stakeholder environment may not be transferable to competitors or future offshore wind market conditions and SR requirements (Puroila & Mäkelä, 2019).

The analysis further entails similar trade-offs to the ones stated in Calabrese et al. (2017) and may not be representative of organizational preference based on the sample size of DMs. A certain degree of simplification and subjectivity through aggregation of both internal and stakeholders' preferences tends to remain in the data and the method due to the topic's nature. Under this consideration, the solutions and implications may not be definitive and require further examination and are unlikely to relieve Cadeler A/S from future SR obligations. This holds especially for the stakeholders' SR screening as it assumes equal levels of stakeholder salience, which will have to examined further with similar MCDM tools as the one applied in this study and may be based on subjective data and translations based on a framework dictionary. In part, stakeholders' preferences may have been misinterpreted or over- or understated. In terms of frameworks and regulation, inefficiencies in regulation and frameworks call for voluntary self-commitments in organizations which can, despite objectification, result in inefficient decision-making and opportunistic behavior. Importantly, the method applied does not provide a true and optimal solution inherent to complex linear programming setups, but rather a suggestive point of departure for adopting materiality-driven SR based on GRI that may not require a clear-cut suggestions present human factors in this approach. In addition, it may ignore potential interdependencies between disclosures, topics, and SDGs though the applied tool follows the "drill-down" approach suggested by GRI.

5.5. Avenues for Future Research

The author concludes with potentially fruitful avenues for future research. To start with, further testing and development of data-driven materiality assessments is required to not only structure quanlitative and relatively weakly standardized procedures, but also taking informative reporting mainstream at a greater level of convenience (Calabrese et al., 2017, 2016). Using GRI metrics for performance evaluation and measurement can shed more light on whether the integration of such a framework renders decision-making more efficient as a result of increased congruity and incentive effects and how well materiality-driven SR enhances frameworks such as the SBSC or the more recent MBSC (Feltham & Xie, 1994; Guix & Font, 2020; Hubbard, 2009; Ikäheimo, Kallunki, Moilanen, & Schiehll, 2018).

Broader application of group-based MCDM techniques could be used or developed to test group behavior in adopting reporting and improving internal resource efficiency in SR integration (George et al., 2016; Wang & Elhag, 2007). This could be supported by introducing importance scoring models for weighting DMs' influence on organizational processes and further econometric analyses on GRI-based metrics, their comparability, and their contribution towards SD. Further investigations on reporting trade-offs between conciseness and comprehensiveness could further increase our understanding on how organizations and investors provide and are given an optimal amount of content that enables efficient decision-making. On the market level, further event studies could increase our knowledge on drivers of GRI materiality and its impact on investors' decision-making and stock price informativeness in European markets (see Grewal et al., 2020; Schiehll & Kolahgar, 2020). Further empirical testing and longitudinal industry-level studies on developments in GRI adoption and its interplay with regulatory practice or on quantitative materiality assessments can broaden our knowledge on whether the latter leads to improvements in organizational sustainability and stock market performance across sectors and markets such as the offshore wind value chain. Finally, the suggested avenues along with the recent scope of coverage should emphasize practical implications for assurance providers to develop more powerful assessment toolkits and complement wide-spread principle-driven assurance standards with tightened ruling to increase SR efficiency and hold reporting organizations more accountable, especially on future reporting against emerging regulation such as the EU Taxonomy. In many ways, the road towards efficient SR and the EU Green Deal targets is long and ambitious and is likely to require abundance in theories of change to navigate open pathways. The author hopes to have aroused some doubts about the apparent representativeness and objectivity of best efforts to construct credible SR and conduct structured materiality assessments.

References

Abernathy, J., Stefaniak, C., Wilkins, A., & Olson, J. (2017). Literature review and research opportunities on credibility of corporate social responsibility reporting. *American Journal of Business*, *32*, 24-41. doi: 10.1108/ajb-04-2016-0013

Aczél, J., & Saaty, T. L. (1983, 3). Procedures for synthesizing ratio judgements. *Journal of Mathematical Psychology*, *27*, 93-102. doi: 10.1016/0022-2496(83)90028-7

Adams, C. A. (2017). Conceptualising the contemporary corporate value creation process. *Accounting, Auditing and Accountability Journal*, *30*, 906-931. doi: 10.1108/AAAJ-04-2016-2529

Adams, C. A., & Frost, G. R. (2008, 12). Integrating sustainability reporting into management practices. *Accounting Forum*, *32*, 288-302. doi: 10.1016/j.accfor.2008.05.002

Adams, C. A., & Larrinaga, C. (2019). Progress: engaging with organisations in pursuit of improved sustainability accounting and performance. *Accounting, Auditing and Accountability Journal*, *32*, 2367-2394. doi: 10.1108/AAAJ-03-2018-3399

Adams, C. A., & McNicholas, P. (2007). Making a difference: Sustainability reporting, accountability and organisational change. *Accounting, Auditing and Accountability Journal*, *20*, 382-402. doi: 10.1108/09513570710748553

Adams, C. A., & Whelan, G. (2009). Conceptualising future change in corporate sustainability reporting. *Accounting, Auditing and Accountability Journal*, *22*, 118-143. doi: 10.1108/09513570910923033

Agrawal, A., & Hockerts, K. (2019). Impact investing: review and research agenda. *Journal of Small Business and Entrepreneurship*, *6331*. doi: 10.1080/08276331.2018.1551457

Alonso, J. A., & Lamata, M. T. (2006). Consistency in the analytic hierarchy process: A new approach. *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems*, *14*, 445-459. doi: 10.1142/S0218488506004114

Alrazi, B., Villiers, C. D., & Staden, C. J. V. (2015). A comprehensive literature review on, and the construction of a framework for, environmental legitimacy, accountability and proactivity. *Journal of Cleaner Production*, *102*, 44-57. doi: 10.1016/j.jclepro.2015.05.022

Alshehhi, A., Nobanee, H., & Khare, N. (2018). The impact of sustainability practices on corporate financial performance: Literature trends and future research potential. *Sustainability (Switzerland)*, *10*. doi: 10.3390/su10020494

Amir, A. Z., & Serafeim, G. (2018). Why and how investors use esg information: Evidence from a global survey. *Financial Analysts Journal*, *74*, 87-103. doi: 10.2469/faj.v74.n3.2

Amran, A., Lee, S. P., & Devi, S. S. (2014). The influence of governance structure and strategic corporate social responsibility toward sustainability reporting quality. *Business Strategy and the Environment*, *23*, 217-235. doi: 10.1002/bse.1767

Anderson, D. R., Anderson, K. E., & Able, T. (2009). Sustainability risk management. *Risk Management*, *12*, 25-38.

Arena, M., & Azzone, G. (2012). A process-based operational framework for sustainability reporting in smes. *Journal of Small Business and Enterprise Development*, *19*, 669-686. doi: 10.1108/14626001211277460

Aull-Hyde, R., Erdogan, S., & Duke, J. M. (2006). An experiment on the consistency of aggregated comparison matrices in ahp. *European Journal of Operational Research*, *171*, 290-295. doi: 10.1016/j.ejor.2004.06.037

Authority, D. B. (2015). *Implementation in denmark of eu directive 2014/95/eu on the disclosure of non-financial information.* Retrieved from http://csrgov.dk/file/557863/implementation-of-eu-directive.pdf

Badia, F., Bracci, E., & Tallaki, M. (2020). Quality and diffusion of social and sustainability reporting in italian public utility companies. *Sustainability (Switzerland)*, *12*, 1-17. doi: 10.3390/su12114525

Balusa, B. C., & Gorai, A. K. (2019, 2). Sensitivity analysis of fuzzy-analytic hierarchical process (fahp) decision-making model in selection of underground metal mining method. *Journal of Sustainable Mining*, *18*, 8-17. doi: 10.1016/j.jsm.2018.10.003

Barman, E. (2015). Of principle and principal: Value plurality in the market of impact investing. *Valuation Studies*, *3*, 9-44. doi: 10.3384/vs.2001-5592.15319

Barney, J. (1991, 3). Firm resources and sustained competitive advantage. *Journal of Management*, *17*, 99-120. Retrieved from http://journals.sagepub.com/doi/10.1177/014920639101700108 doi: 10.1177/014920639101700108

Baron, R. (2014). The evolution of corporate reporting for integrated performance. *Organisation for Economic Co-operation and Development (OECD)*, 1-35.

Battaglia, M., Passetti, E., Bianchi, L., & Frey, M. (2016, 11). Managing for integration: a longitudinal analysis of management control for sustainability. *Journal of Cleaner Production*, *136*, 213-225. doi: 10.1016/j.jclepro.2016.01.108

Baumann-Pauly, D., Wickert, C., Spence, L. J., & Scherer, A. G. (2013). Organizing corporate social responsibility in small and large firms: Size matters. *Journal of Business Ethics*, *115*, 693-705. doi: 10.1007/s10551-013-1827-7

Baumüller, J., & Schaffhauser-Linzatti, M.-M. (2018). In search of materiality for nonfinancial information—reporting requirements of the directive 2014/95/euwesentlichkeit(en) in der nichtfinanziellen berichterstattung – eine analyse der berichtspflichten gemäß richtlinie 2014/95/eu. *NachhaltigkeitsManagementForum | Sustainability Management Forum*, *26*, 101-111. doi: 10.1007/s00550-018-0473-z

Bebbington, J., & Larrinaga, C. (2014). Accounting and sustainable development: An exploration. *Accounting, Organizations and Society*, *39*, 395-413. doi: 10.1016/j.aos.2014.01.003

Bebbington, J., Larrinaga, C., & Moneva, J. M. (2008). Corporate social reporting and reputation risk management. *Accounting, Auditing & Accountability Journal*, *21*, 337-361. doi: 10.1108/09513570810863932

Bebbington, J., Österblom, H., Crona, B., Jouffray, J. B., Larrinaga, C., Russell, S., & Scholtens, B. (2019). Accounting and accountability in the anthropocene. *Accounting, Auditing and Accountability Journal*, *33*, 152-177. doi: 10.1108/AAAJ-11-2018-3745

Bellantuono, N., Pontrandolfo, P., & Scozzi, B. (2016). Capturing the stakeholders ' view in sustainability reporting : A novel approach. *Sustainability (Switzerland)*, *8*, 1-12. doi: 10.3390/su8040379

Beske, F., Haustein, E., & Lorson, P. C. (2020). Materiality analysis in sustainability and integrated reports. *Sustainability Accounting, Management and Policy Journal*, *11*, 162-186. doi: 10.1108/SAMPJ-12-2018-0343

Blowfield, M., & Murray, A. (2008). *Socially responsible investment.* Oxford University Press. doi: 10.1111/j.1468-1331.2012.03886.x

Bocken, N. M., Short, S. W., Rana, P., & Evans, S. (2014). A literature and practice review to develop sustainable business model archetypes. *Journal of Cleaner Production*, *65*, 42-56. Retrieved from http://dx.doi.org/10.1016/j.jclepro.2013.11.039 doi: 10.1016/j.jclepro.2013.11.039

Boiral, O., & Heras-Saizarbitoria, I. (2017, 9). Corporate commitment to biodiversity in mining and forestry: Identifying drivers from gri reports. *Journal of Cleaner Production*, *162*, 153-161. doi: 10.1016/j.jclepro.2017.06.037

Bourguignon, A. (2005). *Management accounting and value creation: The profit and loss of reification* (Vol. 16). doi: 10.1016/j.cpa.2003.03.001

İbrahim Özkan, & Türkşen, I. B. (2014). *Uncertainty and fuzzy decisions.* Springer Netherlands. Retrieved from https://doi.org/10.1007/978-94-017-8691-1_2 doi: 10.1007/978-94-017-8691-1_2

Brown, H. S., de Jong, M., & Levy, D. L. (2009, 4). Building institutions based on information disclosure: lessons from gri's sustainability reporting. *Journal of Cleaner Production*, *17*, 571-580. doi: 10.1016/j.jclepro.2008.12.009

Brown, H. S., Jong, M. D., Lessidrenska, T., & Mossavar-Rahmani, T. (2007). *The rise of the global reporting initiative (gri) as a case of institutional entrepreneurship corporate social responsibility initiative a cooperative project among.* Retrieved from http://www.hks.harvard.edu/m-rcbg/CSRI/

Buhmann, K. (2018). Neglecting the proactive aspect of human rights due diligence? a critical appraisal of the eu's non-financial reporting directive as a pillar one avenue for promoting pillar two action. *Business and Human Rights Journal*, *3*, 23-45. doi: 10.1017/bhj.2017.24

Buniamin, S., Nazli, N., & Ahmad, N. (2015). An integrative perspective of environmental, social and governance (esg) reporting: A conceptual paper. *Proceedings of the International Conference on Accounting Studies (ICAS) 2015*, 9-16. Retrieved from www.icas.my

Buonocore, J. J., Choma, E., Villavicencio, A. H., Spengler, J. D., Koehler, D. A., Evans, J. S., ... Sanchez-Pina, R. (2019). Metrics for the sustainable development goals: renewable energy and transportation. *Palgrave Communications*, *5*, 1-14. Retrieved from http://dx.doi.org/10.1057/s41599-019-0336-4 doi: 10.1057/s41599-019-0336-4

Burritt, R. L., & Schaltegger, S. (2010, 9). Sustainability accounting and reporting: Fad or trend? *Accounting, Auditing & Accountability Journal*, *23*, 829-846. doi: 10.1108/09513571011080144

Busco, C., Frigo, M. L., Quattrone, P., & Riccaboni, A. (2014). Leading practices in integrated reporting. *Strategic Finance*, 23-32.

Cadeler A/S. (2021). *Sustainable development report 2020.* Retrieved from https://www.cadeler.com/media/1603/cadeler-sd-report-2020.pdf

Calabrese, A., Costa, R., Ghiron, N. L., & Menichini, T. (2017). Materiality analysis in sustainability reporting: a method for making it work in practice. *European Journal of Sustainable Development*, *6*, 439-447. doi: 10.14207/ejsd.2017.v6n3p439

Calabrese, A., Costa, R., Ghiron, N. L., & Menichini, T. (2019). Materiality analysis in sustainability reporting: A tool for directing corporate sustainability towards emerging economic, environmental and social opportunities. *Technological and Economic Development of Economy*, *25*, 1016-1038. doi: 10.3846/tede.2019.10550

Calabrese, A., Costa, R., Levialdi, N., & Menichini, T. (2016). A fuzzy analytic hierarchy process method to support materiality assessment in sustainability reporting. *Journal of Cleaner Production*, *121*, 248-264. doi: 10.1016/j.jclepro.2015.12.005

Calabrese, A., Costa, R., & Menichini, T. (2013). Using fuzzy ahp to manage intellectual capital assets: An application to the ict service industry. *Expert Systems with Applications*, *40*, 3747-3755. Retrieved from http://dx.doi.org/10.1016/j.eswa.2012.12.081 doi: 10.1016/j.eswa.2012.12.081

Carroll, A. B. (2009). A history of corporate social responsibility: Concepts and practices. *The Oxford Handbook of Corporate Social Responsibility*. doi: 10.1093/oxfordhb/9780199211593.003.0002

Cerbone, D., & Maroun, W. (2020). Materiality in an integrated reporting setting: Insights using an institutional logics framework. *British Accounting Review*, *52*, 100876. Retrieved from https://doi.org/10.1016/j.bar.2019.100876 doi: 10.1016/j.bar.2019.100876

Chan, H. K., Sun, X., & Chung, S. H. (2019, 10). When should fuzzy analytic hierarchy process be used instead of analytic hierarchy process? *Decision Support Systems*, *125*, 113114. doi: 10.1016/j.dss.2019.113114

Chang, C. W., Wu, C. R., Lin, C. T., & Chen, H. C. (2007, 3). An application of ahp and sensitivity analysis for selecting the best slicing machine. *Computers and Industrial Engineering*, *52*, 296-307. doi: 10.1016/j.cie.2006.11.006

Chang, D. Y. (1996, 12). Applications of the extent analysis method on fuzzy ahp. *European Journal of Operational Research*, *95*, 649-655. doi: 10.1016/0377-2217(95)00300-2

Chen, S., & Fan, J. (2011). Measuring corporate social responsibility based on a fuzzy analytical hierarchy process. *International Journal of Computer Network and Information Security*, *3*, 13-22. doi: 10.5815/ijcnis.2011.05.02

Chen, T. C. T. (2020, 6). Guaranteed-consensus posterior-aggregation fuzzy analytic hierarchy process method. *Neural Computing and Applications*, *32*, 7057-7068. Retrieved from https://doi.org/10.1007/s00521-019-04211-y doi: 10.1007/s00521-019-04211-y

Cheng, B., Ioannou, I., & Serafeim, G. (2014). Corporate social responsibility and access to finance. *Strategic Management Journal*, *35*, 1-23. doi: 10.1002/smj.2131

Christensen, L. T., Morsing, M., & Thyssen, O. (2013). Csr as aspirational talk. *Organization*, *20*, 372-393. doi: 10.1177/1350508413478310

Churet, C., & Eccles, R. G. (2014). Integrated reporting, quality of management, and financial performance. *Journal of Applied Corporate Finance*, *26*, 56-64.

Conway, E. (2019). Quantitative impacts of mandatory integrated reporting. *Journal of Financial Reporting and Accounting*, *17*, 604-634. doi: 10.1108/JFRA-08-2018-0066

Crane, A., Palazzo, G., Spence, L. J., & Matten, D. (2014). Contesting the value of "creating shared value". *California Management Review*, *56*, 130-153. doi: 10.1525/cmr.2014.56.2.130

Crutzen, P. J. (2006). *The "anthropocene"*. Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007/3-540-26590-2_3 doi: 10.1007/3-540-26590-2_3

Davidson, K. M. (2011, 1). Reporting systems for sustainability: What are they measuring? *Social Indicators Research*, *100*, 351-365. doi: 10.1007/s11205-010-9634-3

Deegan, C. (2002). Introduction: The legitimising effect of social and environmental disclosures – a theoretical foundation. *Accounting, Auditing & Accountability Journal*, *15*, 282-311. doi: 10.1108/09513570210435852

Deegan, C. (2005). *Environmental management accounting: an overview and application of the concept.*

Derchi, G. B., Zoni, L., & Dossi, A. (2020, 7). Corporate social responsibility performance, incentives, and learning effects. *Journal of Business Ethics*, *1*, 3. Retrieved from https://doi.org/10.1007/s10551-020-04556-8 doi: 10.1007/s10551-020-04556-8

Donegan, H. A., & Dodd, F. J. (1991). A note on saaty's random indexes. *Mathematical and Computer Modelling*, *15*, 135-137. doi: 10.1016/0895-7177(91)90098-R

Du, S., & Yu, K. (2020). Do corporate social responsibility reports convey value relevant information? evidence from report readability and tone. *Journal of Business Ethics*. Retrieved from https://doi.org/10.1007/s10551-020-04496-3 doi: 10.1007/s10551-020-04496-3

Dubois, D., & Prade, H. (1988). Modelling uncertainty and inductive inference: A survey of recent non-additive probability systems. *Acta Psychologica*, *68*, 53-78. doi: 10.1016/0001-6918(88)90045-5

Dyck, A., Lins, K. V., Roth, L., & Wagner, H. F. (2019, 3). Do institutional investors drive corporate social responsibility? international evidence. *Journal of Financial Economics*, *131*, 693-714. doi: 10.1016/j.jfineco.2018.08.013

Dyllick, T., & Muff, K. (2016). Clarifying the meaning of sustainable business: Introducing a typology from business-as-usual to true business sustainability. *Organization and Environment*, *29*, 156-174. Retrieved from https://doi.org/10.1177/1086026615575176 doi: 10.1177/1086026615575176

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, *60*, 2835-2857. doi: 10.1287/mnsc.2014.1984

Eccles, R. G., Krusz, M. P., & Serafeim, G. (2012). The need for sector-specific materiality and sustainability reporting standards. *Journal of Applied Corporate Finance*, *24*, 65-71.

EU Technical Expert Group on Sustainable Finance. (2019). Taxonomy technical report. , 1-414. Retrieved from https://ec.europa.eu/info/sites/info/files/business_economy_euro/banking_and_finance/documents/190618-sustainable-finance-teg-report-taxonomy_en.pdf%0A

European Commission. (2014). Directive 2014/95/eu of the european parliament and of the council of 22 october 2014 amending directive 2013/34/eu as regards disclosure of non-financial and diversity information by certain large undertakings and groups. *Official Journal of the European Union*, *2016*, 48-119. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0095&from=EN

European Commission. (2020a). *Boosting offshore renewable energy.* Retrieved from https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2096

European Commission. (2020b). *A european green deal.* Retrieved from https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en

Feltham, G. A., & Xie, J. (1994). Performance measure congruity and diversity in multi-task principal / agent relations published by : American accounting association performance measure congruity and diversity in multi-task principal / agent relations. *The Accounting Review*, *69*, 429-453.

Figge, F., Hahn, T., Schaltegger, S., & Wagner, M. (2002). The sustainability

balanced scorecard - linking sustainability management to business strategy. *Business Strategy and the Environment*, *11*, 269-284. doi: 10.1002/bse.339

Frederick, W. C. (1960). The growing concern over business responsibility. *California Management Review*, *2*, 54-61. doi: 10.2307/41165405

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman.

Friedman, M. (1970). *The social responsibility of business is to increase its profits.* Retrieved from https://www.nytimes.com/1970/09/13/archives/a-friedman-doctrine-the-social-responsibility-of-business-is-to.html

Frishkoff, P. (1970). An empirical investigation of the concept of materiality in accounting. *Journal of Accounting Research*, *8*, 116-129. Retrieved from https://www.jstor.org/stable/2674697

Gauthier, T. D. (2001). Detecting trends using spearman's rank correlation coefficient. *Environmental Forensics*, *2*, 359-362. Retrieved from https://www.tandfonline.com/action/journalInformation?journalCode=uenf20 doi: 10.1080/713848278

George, R. A., Siti-Nabiha, A. K., Jalaludin, D., & Abdalla, Y. A. (2016). Barriers to and enablers of sustainability integration in the performance management systems of an oil and gas company. *Journal of Cleaner Production*, *136*, 197-212. doi: 10.1016/j.jclepro.2016.01.097

Germanwatch. (2021). *Reform of the eu non-financial reporting directive: A push towards future-proof reporting obligations.* Retrieved from http://allianceforcorporatetransparency.org/news/landmark-research.html.

Gerwanski, J. (2020). Managers' incentives and disincentives to engage with integrated reporting or why managers might not adopt integrated reporting: an exploratory study in a nascent setting. *Qualitative Research in Accounting and Management*, *17*, 553-587. doi: 10.1108/QRAM-01-2019-0025

Gerwanski, J., Kordsachia, O., & Velte, P. (2019). Determinants of materiality disclosure quality in integrated reporting: Empirical evidence from an international setting. *Business Strategy and the Environment*, *28*, 750-770. Retrieved from https://doi.org/10.1002/bse.2278 doi: 10.1002/bse.2278

Global Reporting Initiative. (2013). *Gri g4: Implementation manual.* Retrieved from https://commdev.org/publications/global-reporting-initiative-g4-sustainability-reporting-guidelines-implementation-manual/

Global Reporting Initiative. (2017). *Mapping g4 to the gri standards.* Retrieved from https://www.globalreporting.org/standards/media/1098/mapping-g4-to-the-gri-standards-disclosures-full-overview.pdf

Global Reporting Initiative. (2020a). *Consolidated set of gri sustainability reporting standards.* Retrieved from https://www.globalreporting.org/how-to-use-the-gri-standards/resource-center/

Global Reporting Initiative. (2020b). *Linking the sdgs and the gri standards.* Retrieved from https://www.globalreporting.org/public-policy-partnerships/sustainable-development/integrating-sdgs-into-sustainability-reporting/

Global Reporting Initiative, & University of Stellenbosch Business School. (2020). *Sustainability reporting policy: Global trends in disclosure as the esg agenda goes mainstream.* Retrieved from https://www.carrotsandsticks.net/media/zirbzabv/carrots-and-sticks-2020-interactive.pdf

Gond, J. P., Grubnic, S., Herzig, C., & Moon, J. (2012). Configuring management control systems: Theorizing the integration of strategy and sustainability. *Management Accounting Research*, *23*, 205-223. Retrieved from http://dx.doi.org/10.1016/j.mar.2012.06.003 doi: 10.1016/j.mar.2012.06.003

Govindan, K., Khodaverdi, R., & Jafarian, A. (2013, 5). A fuzzy multi criteria approach for measuring sustainability performance of a supplier based on triple bottom line approach. *Journal of Cleaner Production*, *47*, 345-354. doi: 10.1016/j.jclepro.2012.04.014

Gray, G. L., Turner, J. L., Coram, P. J., & Mock, T. J. (2011). Perceptions and misperceptions regarding the unqualified auditor's report by financial statement preparers, users, and auditors. *Accounting Horizons*, *25*, 659-684. doi: 10.2308/acch-50060

Gray, R. (2006). Social, environmental and sustainability reporting and organisational value creation?: Whose value? whose creation? *Accounting, Auditing and Accountability Journal*, *19*, 793-819. doi: 10.1108/09513570610709872

Green, P. E. J. (2015). *Introduction to risk management principles.* Elsevier Science & Technology.

Grewal, J., Hauptmann, C., & Serafeim, G. (2020). Material sustainability information and stock price informativeness. *Journal of Business Ethics*. doi: 10.1007/s10551-020-04451-2

Guidry, R. P., & Patten, D. M. (2010). Market reactions to the first-time issuance of corporate sustainability reports: Evidence that quality matters. *Sustainability Accounting, Management and Policy Journal*, *1*, 33-50. doi: 10.1108/20408021011059214

Guix, M., & Font, X. (2020). The materiality balanced scorecard: A framework for stakeholder-led integration of sustainable hospitality management and reporting. *International Journal of Hospitality Management*, *91*, 102634. Retrieved from https://doi.org/10.1016/j.ijhm.2020.102634 doi: 10.1016/j.ijhm.2020.102634

Hallstedt, S., Ny, H., Robèrt, K. H., & Broman, G. (2010, 5). An approach to assessing sustainability integration in strategic decision systems for product development. *Journal of Cleaner Production*, *18*, 703-712. doi: 10.1016/j.jclepro.2009.12.017

Herremans, I. M., & Nazari, J. A. (2016). Sustainability reporting driving forces and management control systems. *Journal of Management Accounting Research*, *28*, 103-124. doi: 10.2308/jmar-51470

Hsu, C. W., Lee, W. H., & Chao, W. C. (2013). Materiality analysis model in sustainability reporting: A case study at lite-on technology corporation. *Journal of Cleaner Production*, *57*, 142-151. Retrieved from http://dx.doi.org/10.1016/j.jclepro.2013.05.040 doi: 10.1016/j.jclepro.2013.05.040

Hubbard, G. (2009). Measuring organizational performance: Beyond the triple bottom line. *Business Strategy and the Environment*, *18*, 177-191. doi: 10.1002/bse.564

Ikäheimo, S., Kallunki, J. P., Moilanen, S., & Schiehll, E. (2018). Do white-collar employee incentives improve firm profitability? *Journal of Management Accounting Research*, *30*, 95-115. doi: 10.2308/JMAR-51902

International Integrated Reporting Council. (2013). *The international <ir> framework.* Retrieved from https://integratedreporting.org/wp-content/uploads/2015/03/13-12-08-THE-INTERNATIONAL-IR-FRAMEWORK-2-1.pdf

Ioannou, I., & Serafeim, G. (2012). What drives corporate social performance the role of nation-level institutions. *Journal of International Business Studies*, *43*, 834-864. doi: 10.1057/jibs.2012.26

Jackson, E. T. (2013). Interrogating the theory of change: evaluating impact investing where it matters most. *Journal of Sustainable Finance and Investment*, *3*, 95-110. doi: 10.1080/20430795.2013.776257

Jebe, R. (2019). The convergence of financial and esg materiality: Taking sustainability mainstream. *American Business Law Journal*, *56*, 645-702. doi: 10.1111/ablj.12148

Jensen, J. C., & Berg, N. (2012). Determinants of traditional sustainability reporting versus integrated reporting. an institutionalist approach. *Business Strategy and the Environment*, *21*, 299-316. doi: 10.1002/bse.740

Jespersen, K., & Olmsted, P. (2019). Strategic approaches to sustainability reporting.

Johansen, T. R. (2016). Eu regulation of corporate social and environmental reporting. *Social and Environmental Accountability Journal*, *36*, 1-9. Retrieved from http://dx.doi.org/10.1080/0969160X.2016.1148948 doi: 10.1080/0969160X.2016.1148948

Jones, T. M. (1980). Corporate social responsibility revisited, redefined. *California Management Review*, *22*, 59-67. doi: 10.2307/41164877

Jong, M. D. D., Harkink, K. M., & Barth, S. (2018). Making green stuff? effects of corporate greenwashing on consumers. *Journal of Business and Technical Communication*, *32*, 77-112. doi: 10.1177/1050651917729863

Junior, R. M., Best, P. J., & Cotter, J. (2014). Sustainability reporting and assurance: A historical analysis on a world-wide phenomenon. *Journal of Business Ethics*, *120*, 1-11. doi: 10.1007/s10551-013-1637-y

Kaufmann, A., & Gupta, M. M. (1991). *Introduction to fuzzy arithmetic: Theory and applications*. Van Nostrand Reinhold.

Kerr, J., Rouse, P., & de Villiers, C. (2015). Sustainability reporting inte-

grated into management control systems. *Pacific Accounting Review*, *27*, 189-207. doi: 10.1108/PAR-08-2012-0034

Khan, H. Z., Bose, S., Mollik, A. T., & Harun, H. (2020). "green washing" or "authentic effort"? an empirical investigation of the quality of sustainability reporting by banks. *Accounting, Auditing and Accountability Journal*, *34*, 338-369. doi: 10.1108/AAAJ-01-2018-3330

Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate sustainability: First evidence on materiality. *Accounting Review*, *91*, 1697-1724. doi: 10.2308/accr-51383

Kiesnere, A. L., & Baumgartner, R. J. (2019). Sustainability management emergence and integration on different management levels in smaller large-sized companies in austria. *Corporate Social Responsibility and Environmental Management*, *26*, 1607-1626. Retrieved from https://doi.org/10.1002/csr.1854 doi: 10.1002/csr.1854

Koep, L. (2017). Tensions in aspirational csr communication-a longitudinal investigation of csr reporting. *Sustainability (Switzerland)*, *9*. doi: 10.3390/su9122202

KPMG. (2016). *Carrots & sticks: Global trends in sustainability reporting regulation and policy.* Retrieved from https://assets.kpmg/content/dam/kpmg/pdf/2016/05/carrots-and-sticks-may-2016.pdf

KPMG. (2020). *The time has come: The kpmg survey of sustainability reporting 2020.* Retrieved from https://assets.kpmg/content/dam/kpmg/xx/pdf/2020/11/the-time-has-come.pdf

Krejčí, J. (2018). *Pairwise comparison matrices and their fuzzy extension. multi-criteria decision making with a new fuzzy approach*.

La Torre, M., Sabelfeld, S., Blomkvist, M., & Dumay, J. (2020). Rebuilding trust: sustainability and non-financial reporting and the european union regulation. *Meditari Accountancy Research*, *28*, 701-725. doi: 10.1108/MEDAR-06-2020-0914

La Torre, M., Sabelfeld, S., Blomkvist, M., Tarquinio, L., & Dumay, J. (2018). Harmonising non-financial reporting regulation in europe: Practical forces and projections for future research. *Meditari Accountancy Research*, *26*, 598-621. doi: 10.1108/MEDAR-02-2018-0290

Laarhoven, P. J. M. V., & Pedryczt, W. (1983). *A fuzzy extension of saaty's priority theory* (Vol. 11).

Lee, A. R. (1995). *Application of modified fuzzy ahp method to analyze bolting sequence of structural joints* (Unpublished doctoral dissertation). Lehigh University Betlehem, PA, United States.

Lee, S. H. (2010, 7). Using fuzzy ahp to develop intellectual capital evaluation model for assessing their performance contribution in a university. *Expert Systems with Applications*, *37*, 4941-4947. doi: 10.1016/j.eswa.2009.12.020

Lucarelli, C., Mazzoli, C., Rancan, M., & Severini, S. (2020). Classification of sustainable activities: Eu taxonomy and scientific literature. *Sustainability (Switzerland)*, *12*. doi: 10.3390/su12166460

Ma, Y., Zhang, Q., Yin, Q., & Wang, B. (2019). The influence of top managers on environmental information disclosure: The moderating effect of company's environmental performance. *International Journal of Environmental Research and Public Health*, *16*. doi: 10.3390/ijerph16071167

Maas, K., Schaltegger, S., & Crutzen, N. (2016, 11). Reprint of advancing the integration of corporate sustainability measurement, management and reporting. *Journal of Cleaner Production*, *136*, 1-4. doi: 10.1016/j.jclepro.2016.08.055

Machado, B. A. A., Dias, L. C. P., & Fonseca, A. (2021). Transparency of materiality analysis in gri-based sustainability reports. *Corporate Social Responsibility and Environmental Management*, *28*, 570-580. doi: 10.1002/csr.2066

Maguire, M. (2010). The future of corporate social responsibility reporting. *The Fredercik S. Pardee Center for the Study of the Longer-Range Future, Boston University*. Retrieved from http://www.bu.edu/pardee/files/2011/01/PardeeIIB-019-Jan-2011.pdf

Majoch, A. A., Hoepner, A. G., & Hebb, T. (2017). Sources of stakeholder salience in the responsible investment movement: Why do investors sign the principles for responsible investment? *Journal of Business Ethics*, *140*, 723-741. doi: 10.1007/s10551-016-3057-2

Majumdar, A., Tiwari, M. K., Agarwal, A., & Prajapat, K. (2021, 4). A new case of rank reversal in analytic hierarchy process due to aggregation of cost and benefit criteria. *Operations Research Perspectives*, *8*, 100185. doi: 10.1016/j.orp.2021.100185

Mardani, A., Jusoh, A., & Zavadskas, E. K. (2015, 5). *Fuzzy multiple criteria decision-making techniques and applications - two decades review from 1994 to 2014* (Vol. 42). Elsevier Ltd. doi: 10.1016/j.eswa.2015.01.003

Mattiussi, A., Rosano, M., & Simeoni, P. (2014, 1). A decision support system for sustainable energy supply combining multi-objective and multi-attribute analysis: An australian case study. *Decision Support Systems*, *57*, 150-159. doi: 10.1016/j.dss.2013.08.013

Matuszyk, I., & Rymkiewicz, B. (2018). Integrated reporting and sustainable development reporting-comparison of guidelines iirc and gri g4. *European Journal of Management and Economics*, *6*, 31-43. Retrieved from www.ceejme.euwww.wsb.pl/wroclaw/ceejmehttp://dx.doi.org/10.29015/ceejme.542. doi: 10.29015/ceejme.542

Mauro, S. G., Cinquini, L., Simonini, E., & Tenucci, A. (2020). Moving from social and sustainability reporting to integrated reporting: Exploring the potential of italian public-funded universities' reports. *Sustainability (Switzerland)*, *12*, 3172. doi: 10.3390/SU12083172

Melloni, G., Caglio, A., & Perego, P. (2016). *Saying more with less? disclosure conciseness, completeness and balance in integrated reports.*

Michelon, G., Pilonato, S., & Ricceri, F. (2015, 12). Csr reporting practices and the quality of disclosure: An empirical analysis. *Critical Perspectives on Accounting*, *33*, 59-78. doi: 10.1016/j.cpa.2014.10.003

Mies, A., & Neergaard, P. (2020). Quality of csr reporting: Mandatory or voluntary reporting? *Developments in Corporate Governance and Responsibility*, *15*, 205-234. doi: 10.1108/S2043-052320200000015012

Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review*, *22*, 853-886. doi: 10.5465/AMR.1997.9711022105

Moon, J. (2005). *United kingdom: An explicit model of business-society relations.* Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007/3-540-26960-6_5 doi: 10.1007/3-540-26960-6_5

Morgan, A. (2014). Empirically analyzing the "five percent rule of materiality" in financial reporting decisions. *Senior Honors Projects*, *27*, 1-25. Retrieved from https://collected.jcu.edu/cgi/viewcontent.cgi?article=1026&context=honorspapers

Morioka, S. N., Evans, S., & Carvalho, M. M. D. (2016, 1). Sustainable business model innovation: Exploring evidences in sustainability reporting. In (Vol. 40, p. 659-667). Elsevier B.V. doi: 10.1016/j.procir.2016.01.151

Morrison-Saunders, A., & Therivel, R. (2006, 9). Sustainability integration and assessment. *Journal of Environmental Assessment Policy and Management*, *8*, 281-298. doi: 10.1142/S1464333206002529

Morsing, M., & Thyssen, C. (2003). Corporate values and responsbility - the case of denmark.

Moseñe, J. A., Burritt, R. L., Sanagustín, M. V., Moneva, J. M., & Tingey-Holyoak, J. (2013, 2). Environmental reporting in the spanish wind energy sector: An institutional view. *Journal of Cleaner Production*, *40*, 199-211. doi: 10.1016/j.jclepro.2012.08.023

Nigri, G., & Baldo, M. D. (2018). Sustainability reporting and performance measurement systems: How do small- and medium- sized benefit corporations manage integration? *Sustainability (Switzerland)*, *10*. Retrieved from www.mdpi.com/journal/sustainability doi: 10.3390/su10124499

O'Dwyer, B., & Unerman, J. (2020). Shifting the focus of sustainability accounting from impacts to risks and dependencies: researching the transformative potential of tcfd reporting. *Accounting, Auditing and Accountability Journal*, *33*, 1113-1141. doi: 10.1108/AAAJ-02-2020-4445

Oikonomou, I., Yin, C., & Zhao, L. (2020). Investment horizon and corporate social performance: the virtuous circle of long-term institutional ownership and responsible firm conduct. *European Journal of Finance*, *26*, 14-40. doi: 10.1080/1351847X.2019.1660197

Okhuysen, G. A., & Bechky, B. A. (2009). Coordination in organizations: An integrative perspective. *The Academy of Management Annals*, *3*, 463-502. doi: 10.1080/19416520903047533

Orlitzky, M. (2013). Corporate social responsibility, noise, and stock market volatility. *Academy of Management Perspectives*, *27*, 238-254. doi: 10.5465/amp.2012.0097

Orsato, R. J. (2006). Competitive environmental strategies: When does it pay to be green? *California Management Review*, *48*, 127-143.

Orsato, R. J. (2009). *What are sustinability strategies?* Houndmills UK: Palgrave Macmillan.

Pandey, M. (2012). *New aggregation operator for triangular fuzzy numbers based on the geometric means of the slopes of the l-and r-membership functions* (Vol. 2). Retrieved from www.ijctonline.com

Petcharat, N., & Zaman, M. (2019). Sustainability reporting and integrated reporting perspectives of thai-listed companies. *Journal of Financial Reporting and Accounting*, *17*, 671-694. doi: 10.1108/JFRA-09-2018-0073

Pizzi, S., Rosati, F., & Venturelli, A. (2021). The determinants of business contribution to the 2030 agenda: Introducing the sdg reporting score. *Business Strategy and the Environment*, *30*, 404-421. Retrieved from https://doi.org/10.1002/bse.2628 doi: 10.1002/bse.2628

Pizzi, S., Venturelli, A., & Caputo, F. (2020). The "comply-or-explain" principle in directive 95/2014/eu. a rhetorical analysis of italian pies. *Sustainability Accounting, Management and Policy Journal*, *12*, 30-50. Retrieved from https://www.emerald.com/insight/2040-8021.htm doi: 10.1108/SAMPJ-07-2019-0254

Porter, M. E., & Kramer, M. R. (2011). Creating shared value. *Harvard Business Review*, 1-17.

Power, M. (2009). The risk management of nothing. *Accounting, Organizations and Society*, *34*, 849-855. Retrieved from http://dx.doi.org/10.1016/j.aos.2009.06.001 doi: 10.1016/j.aos.2009.06.001

Puroila, J., Kujala, J., & Mäkelä, H. (2016, 11). Reframing materiality in sustainability reporting. *Academy of Management Proceedings*, *2016*, 12697. Retrieved from https://journals.aom.org/doi/abs/10.5465/ambpp.2016.12697abstract doi: 10.5465/ambpp.2016.12697abstract

Puroila, J., & Mäkelä, H. (2018, 8). Materiality in sustainability reporting: An illusion of consensus and objectivity? *Academy of Management Proceedings*, *2018*, 16174. Retrieved from https://journals.aom.org/doi/abs/10.5465/AMBPP.2018.16174abstract doi: 10.5465/ambpp.2018.16174abstract

Puroila, J., & Mäkelä, H. (2019). Matter of opinion: Exploring the socio-political nature of materiality disclosures in sustainability reporting. *Accounting, Auditing and Accountability Journal*, *32*, 1043-1072. doi: 10.1108/AAAJ-11-2016-2788

Quatrini, S. (2021). Challenges and opportunities to scale up sustainable finance after the covid-19 crisis: Lessons and promising innovations from science and practice. *Ecosystem Services*, *48*, 101240. Retrieved from https://doi.org/10.1016/j.ecoser.2020.101240 doi: 10.1016/j.ecoser.2020.101240

Rasche, A., Gwozdz, W., Larsen, M. L., & Moon, J. (2020). Which firms leave multi-stakeholder initiatives? an analysis of delistings from the united nations global compact. *Regulation and Governance*. doi: 10.1111/rego.12322

Reimsbach, D., Schiemann, F., Hahn, R., & Schmiedchen, E. (2020). In the eyes of the beholder: Experimental evidence on the contested nature of materiality in sustainability reporting. *Organization and Environment*, *33*, 624-651. Retrieved from https://doi.org/10.1177/1086026619875436 doi: 10.1177/1086026619875436

Reynolds, M., & Yuthas, K. (2008). Moral discourse and corporate social responsibility reporting. *Journal of Business Ethics*, *78*, 47-64. doi: 10.1007/s10551-006-9316-x

Richardson, J. (2008). The business model: an integrative framework for strategy execution. *Strategic Change*, *17*, 133-144. doi: 10.1002/jsc.821

Robinson, J. (2004). Squaring the circle? some thoughts on the idea of sustainable development. *Ecological Economics*, *48*, 369-384. doi: 10.1016/j.ecolecon.2003.10.017

Rodríguez-Gutiérrez, P., Guerrero-Baena, M. D., Luque-Vílchez, M., & Castilla-Polo, F. (2021). An approach to using the best-worst method for supporting sustainability reporting decision-making in smes. *Journal of Environmental Planning and Management*, *0*, 1-24. Retrieved from https://doi.org/10.1080/09640568.2021.1876003 doi: 10.1080/09640568.2021.1876003

Rodríguez-Olalla, A., & Avilés-Palacios, C. (2017). Integrating sustainability in organisations: An activity-based sustainability model. *Sustainability (Switzerland)*, *9*, 1-18. doi: 10.3390/su9061072

Rodrigues, M., & Franco, M. (2019). The corporate sustainability strategy in organisations: A systematic review and future directions. *Sustainability (Switzerland)*, *11*. doi: 10.3390/su11226214

Rubino, F., & Napoli, F. (2020). What impact does corporate governance have on corporate environmental performances? an empirical study of italian listed firms. *Sustainability (Switzerland)*, *12*, 1-21. doi: 10.3390/su12145742

Rupley, K. H., Brown, D., & Marshall, S. (2017, 10). Evolution of corporate reporting: From stand-alone corporate social responsibility reporting to integrated reporting. *Research in Accounting Regulation*, *29*, 172-176. doi: 10.1016/j.racreg.2017.09.010

Saaty, T. L. (1978, 9). Modeling unstructured decision problems - the theory of analytical hierarchies. *Mathematics and Computers in Simulation*, *20*, 147-158. doi: 10.1016/0378-4754(78)90064-2

Saaty, T. L., & Tran, L. T. (2007, 10). On the invalidity of fuzzifying numerical judgments in the analytic hierarchy process. *Mathematical and Computer Modelling*, *46*, 962-975. doi: 10.1016/j.mcm.2007.03.022

Saenz, C. (2019). Creating shared value using materiality analysis: Strategies from the mining industry. *Corporate Social Responsibility and Environmental Management*, *26*, 1351-1360. doi: 10.1002/csr.1751

Safari, M., & Areeb, A. (2020). A qualitative analysis of gri principles for defining sustainability report quality: an australian case from the preparers' perspective. *Accounting Forum*, *44*, 344-375. Retrieved from https://doi.org/10.1080/01559982.2020.1736759 doi: 10.1080/01559982.2020.1736759

Schaltegger, S., Burritt, R., & Petersen, H. (2017). *An introduction to corporate environmental management*. doi: 10.4324/9781351281447

Schiehll, E., & Kolahgar, S. (2020). Financial materiality in the informativeness of sustainability reporting. *Business Strategy and the Environment*, *2017*, 1-16. doi: 10.1002/bse.2657

Schläpfer, F., & Fischoff, B. (2010). When are preferences consistent? the effects of task familiarity and contextual cues on revealed and stated preferences. *Working Paper No. 1007*. Retrieved from http://hdl.handle.net/10419/76190

Scholtens, B., & Sievänen, R. (2013). Drivers of socially responsible investing: A case study of four nordic countries. *Journal of Business Ethics*, *115*, 605-616. doi: 10.1007/s10551-012-1410-7

Schwartz, M. S., & Carroll, A. B. (2003). Corporate social responsibility: A three-domain approach. *Business Ethics Quarterly*, *13*, 503-530.

Serafeim, G., Park, D., Freiberg, D., & Zochowski, R. (2020). Corporate environmental impact: Measurement, data and information. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3565533

Shift & Mazars. (2015). Un guiding principles reporting framework. , 1-114. Retrieved from https://shiftproject.org/wp-content/uploads/2015/02/UNGPReportingFramework_withguidance2017.pdf

Skouloudis, A., Evangelinos, K., & Kourmousis, F. (2009). Development of an evaluation methodology for triple bottom line reports using international standards on reporting. *Environmental Management*, *44*, 298-311. doi: 10.1007/s00267-009-9305-9

Sonnerfeldt, A. L. L., & Pontoppidan, C. A. (2020). The challenges of assurance on non-financial reporting. *Accounting, Economics and Law: A Convivium*, *10*, 731-757. doi: 10.1515/ael-2018-0050

Stacchezzini, R., Melloni, G., & Lai, A. (2016). Sustainability management and reporting: the role of integrated reporting for communicating corporate sustainability management. *Journal of Cleaner Production*, *136*, 102-110. Retrieved from http://dx.doi.org/10.1016/j.jclepro.2016.01.109 doi: 10.1016/j.jclepro.2016.01.109

Stolowy, H., & Paugam, L. (2018). The expansion of non-financial reporting: an exploratory study. *Accounting and Business Research*, *48*, 525-548. Retrieved from https://doi.org/10.1080/00014788.2018.1470141 doi: 10.1080/00014788.2018.1470141

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-15*, 116-132. doi: 10.1109/TSMC.1985.6313399

Talbot, D., & Boiral, O. (2018). Ghg reporting and impression management: An assessment of sustainability reports from the energy sector. *Journal of Business Ethics*, *147*, 367-383. Retrieved from

https://doi.org/10.1007/s10551-015-2979-4 doi: 10.1007/s10551-015-2979-4

Thoradeniya, P., Lee, J., Tan, R., & Ferreira, A. (2015). Sustainability reporting and the theory of planned behaviour. *Accounting, Auditing and Accountability Journal*, *28*, 1099-1137. doi: 10.1108/AAAJ-08-2013-1449

Torelli, R., Balluchi, F., & Furlotti, K. (2020). The materiality assessment and stakeholder engagement: A content analysis of sustainability reports. *Corporate Social Responsibility and Environmental Management*, *27*, 470-484. doi: 10.1002/csr.1813

Traxler, A. A., Schrack, D., & Greiling, D. (2020, 12). *Sustainability reporting and management control – a systematic exploratory literature review* (Vol. 276). Elsevier Ltd. doi: 10.1016/j.jclepro.2020.122725

Tsai, H. Y., Chang, C. W., & Lin, H. L. (2010, 8). Fuzzy hierarchy sensitive with delphi method to evaluate hospital organization performance. *Expert Systems with Applications*, *37*, 5533-5541. doi: 10.1016/j.eswa.2010.02.099

Tschopp, D., & Huefner, R. J. (2015). Comparing the evolution of csr reporting to that of financial reporting. *Journal of Business Ethics*, *127*, 565-577. doi: 10.1007/s10551-014-2054-6

Tseng, M. L., & Lin, Y. H. (2008). Selection of competitive advantages in tqm implementation using fuzzy ahp and sensitivity analysis. *Asia Pacific Management Review*, *13*, 583-599. doi: 10.6126/APMR.2008.13.3.03

Unerman, J., & Zappettini, F. (2014). Incorporating materiality considerations into analyses of absence from sustainability reporting. *Social and Environmental Accountability Journal*, *34*, 172-186. Retrieved from http://dx.doi.org/10.1080/0969160X.2014.965262 doi: 10.1080/0969160X.2014.965262

United Nations. (2015a). *The 2030 agenda for sustainable development.* Retrieved from https://sustainabledevelopment.un.org/content/documents/21252030AgendaforSustainableDevelopmentweb.pdf

United Nations. (2015b). *Global indicator framework for the sustainable development goals and targets of the 2030 agenda for sustainable development.* Retrieved from https://unstats.un.org/sdgs/indicators/GlobalIndicatorFrameworkafter2019refinement_Eng.pdf%0Ahttps://unstats.un.org/sdgs/indicators/GlobalIndicatorFramework_A.RES.71.313Annex.pdf

Velasquez, M., & Hester, P. T. (2013). A comparative analysis of multicriteria decision-making methods. *International Journals of Operations Research*, *10*, 56-66. doi: 10.1007/s13748-016-0093-1

Venturelli, A., Luison, C., Badalotti, G., Bodo, R., Caputo, F., Corvino, A., ... Vermiglio, C. (2020). *The sdgs in the reports of the italian companies* (Vol. 16). Retrieved from https://fair.unifg.it/handle/11369/385596#.XxCICekdbZs.mendeley

Vermeulen, F., & Barkema, H. (2002). Pace, rhythm, and scope: Process dependence in building a profitable multinational corporation. *Strategic Management Journal*, *23*, 637-653. doi: 10.1002/smj.243

Villiers, C. D., & Maroun, W. (2017). *Introduction to sustainability accounting and integrated reporting.* doi: 10.4324/9781315108032-1

Vinnari, E., & Laine, M. (2013). Just a passing fad?: The diffusion and decline of environmental reporting in the finnish water sector. *Accounting, Auditing and Accountability Journal*, *26*, 1107-1134. doi: 10.1108/AAAJ-04-2012-01002

Vogel, D. (2006). *The market for virtue: The potential and limits of corporate social responsibility* (2nd ed.). Brookings Institution Press. Retrieved from http://www.jstor.org/stable/10.7864/j.ctt6wpg2c

Wang, Y. M., & Elhag, T. M. (2006). On the normalization of interval and fuzzy weights. *Fuzzy Sets and Systems*, *157*, 2456-2471. doi: 10.1016/j.fss.2006.06.008

Wang, Y. M., & Elhag, T. M. (2007, 8). A fuzzy group decision making approach for bridge risk assessment. *Computers and Industrial Engineering*, *53*, 137-148. doi: 10.1016/j.cie.2007.04.009

Wang, Y. M., Luo, Y., & Hua, Z. (2008). On the extent analysis method for fuzzy ahp and its applications. *European Journal of Operational Research*, *186*, 735-747. doi: 10.1016/j.ejor.2007.01.050

Weber, O. (2016). *Finance and sustainability*. doi: 10.1007/978-94-017-7242-6_10

Whitehead, J. (2017). Prioritizing sustainability indicators: Using materiality analysis to guide sustainability assessment and strategy. *Business Strategy and the Environment*, *26*, 399-412. doi: 10.1002/bse.1928

Wind, Y., & Saaty, T. L. (1980). Marketing applications of the analytic hierarchy process. *Management Science*, *26*, 641-658. doi: 10.1287/mnsc.26.7.641

Winkler, P., Etter, M., & Castelló, I. (2020). Vicious and virtuous circles of aspirational talk: From self-persuasive to agonistic csr rhetoric. *Business and Society*, *59*, 98-128. doi: 10.1177/0007650319825758

Witjes, S., Vermeulen, W. J., & Cramer, J. M. (2017, 12). Assessing corporate sustainability integration for corporate self-reflection. *Resources, Conservation and Recycling*, *127*, 132-147. doi: 10.1016/j.resconrec.2017.08.026

Wren, D. A. (2005). *The history of management thought* (5th Edition ed.). John Wiley & Sons, Inc.

Wu, S. R., Shao, C., & Chen, J. (2018). Approaches on the screening methods for materiality in sustainability reporting. *Sustainability (Switzerland)*, *10*, 14-16. doi: 10.3390/su10093233

Zhou, Y. (2011, 11). Materiality approach in sustainability reporting: applications, dilemmas, and challenges. In (p. 548). MDPI. Retrieved from https://doi.org/10.3390/wsf-00548 doi: 10.3390/wsf-00548

Zoomers, E. B. A., & Otsuki, K. (2017). Addressing the impacts of large-scale land investments: Re-engaging with livelihood research. *Geoforum*, *83*, 164-171. doi: 10.1016/j.geoforum.2017.01.009

Junior Management Science 7(3) (2022) 731-755

Junior Management Science

journal homepage: www.jums.academy

Carbon Pricing: A Comparison between Germany and the United Kingdom

Karina Temoche González

Technische Universität München

Abstract

Climate change is a global problem that almost every country – 191 parties had signed the Paris Agreement - has committed to undertake. The European Union (EU) has been one of the pioneers in implementing policies that tackle greenhouse gas emissions (GHG). In 2005, the European Emissions Trading Scheme (EU ETS) was launched as the first carbon market. Despite the EU ETS evolving throughout the years, the United Kingdom (UK) implemented an additional policy. In 2013, the UK introduced a Carbon Price Floor (CPF). This paper examines the impact of carbon pricing on GHG emissions during phase III of the EU ETS (2017-2020) in Germany and the UK. Electricity generated by nuclear and renewable sources are considered in the analysis. There are two research questions. First, is the impact of carbon pricing in these two countries, measured by using an Ordinary Least Squares (OLS) model for panel data. The results show that the UK has been more successful in reducing GHG emissions because of the CPF implementation. Second, whether the Market Stability Reserve (MSR) – a policy within the EU ETS – acted as a Carbon Price Floor (CPF) for Germany. Using a model of Differences in Differences (DD), this paper showed that the MSR significantly reduced the CO2 emissions of Germany.

Keywords: Carbon price; EU ETS; CO2 emissions; carbon price floor; market stability reserve; differences in differences.

1. Introduction

The European Union (EU) recognizes climate change as one of the most significant economic, social, and environmental challenges that the world faces (Bruggeman & Gonenc, 2013). The climate goals to reduce greenhouse gases (GHG) emissions are led by international commitments such as the 1997 Kyoto Protocol and the 2015 Paris Agreement, the last was ratified by 191 parties (including the EU) (The United Nations, 2021). Before the 2015 Paris Agreement, the EU implemented policies in various areas to tackle climate change and fulfill its GHG emissions reduction targets of 20% by 2020, 40% by 2030, and climate neutrality by 2050 (European Environment Agency, 2020). One of the tools to combat climate change is the European Emissions Trading System (EU ETS) which is the largest and the world's first carbon market since 2005 (European Commission, 2021a). At the time of its creation and until 2020, the scheme included the United Kingdom (UK) as a participant. The EU ETS operates in phases, whose align progressively to the EU climate policy objectives. Phase 1 (2005 - 2007) was a pilot phase, phase 2 (2008 - 2012) comprised the first commitment period of the Kyoto Protocol, and phase 3 (2013 - 2020) comprised the second commitment period of the Kyoto Protocol (European Commission, 2021a).

In April 2013, at the beginning of the third phase of the EU ETS, the UK Government introduced a Carbon Price Floor (CPF) as a complementary measure to the EU ETS (Hirst, 2018). The British Government launched the UK CPF on top of the EU ETS to promote long-term investments in clean technologies. According to the UK Government, the price of the European Allowances (EUAs) was not high enough to support these risky investments, which are necessary to accomplish British environmental goals (Hirst, 2018). Since its implementation, the rate of this tax has oscillated between £5 - £18. At the same time, the EU Commission introduced some reforms to strengthen the EUA price. Considering that taxes are paid per ton of CO_2, the marginal cost of high polluter fossil-fuel power plants has increased considerably more compared to the less polluting ones. As a consequence, most countries in Europe have a cleaner electricity. However, the mixes of electricity have changed differently.

By 2020, Germany and the UK generated 44.9% and 42.3% of their electricity from renewable sources (i.e. hydro, solar, wind, and other renewables) (Our World in Data,

DOI: https://doi.org/10.5282/jums/v7i3pp731-755

2021). When nuclear energy is included, 56.2% and 59.3% of German and British electricity, respectively, is generated by clean sources (Our World in Data, 2021). Nevertheless, electricity generated by fossil-fuel sources differs significantly. In Germany, 23.7% of electricity was generated by coal (hard coal plus lignite), in the UK only 1.7% (Our World in Data, 2021). Studies carried by Gugler, Haxhimusa, and Liebensteiner (2021) and Wilson and Staffell (2018) have compared the impact of carbon pricing between these two countries. Both authors agreed that the UK CPF has been more effective in reducing CO_2 emissions. However, both authors have missed the consideration of two facts: 1) the nuclear policy, and 2) the Market Stability Reserve (MSR).

In 2011, Germany officially announced that the country will shut down all its nuclear power plants by 2022 (World Nuclear Association, 2021a). Conversely, the UK supports nuclear energy and recognizes it as fundamental to fulfill its environmental goals (World Nuclear Association, 2021b). Between these two countries, only Germany is closing nuclear power plants. Could this factor influence the success of the CPF in the UK when compared to Germany? Considering that only one of them has to replace a reliable electricity source that represents more than 10% of its electricity mix, nuclear phase-out may be a factor. Especially because fossil fuels are the other reliable source available, the only capable to substitute nuclear. This document will include electricity generated by nuclear energy as an explanatory variable for the CO_2 emissions. If the variable is found significant, the model will produce a robust estimator of the relationship between the carbon price and CO_2 emissions, as well. The latter is the first goal of this document.

On the other hand, the implementation of the MSR in 2019 stabilized the price of the EUA. For instance, during the Covid-19 crisis, the price of the EUA fell to 16 € /ton, but it recovered its previous value after four months. On top of that, since the MSR was implemented, the EUA had experienced an uptrend. Edenhofer et al. (2017) and Schmidt (2020) concluded that the MSR reform was less effective than a CPF to promote decarbonization. Nevertheless, these authors did not compare its effectiveness with the UK CPF. Could the MSR act as a CPF? This paper will test the behavior of the CO_2 emissions in Germany after its implementation in 2019.

The Ordinary Least Squares (OLS) model for panel data will be used to test the influence of carbon pricing in tackling the CO_2 emissions of the UK and Germany. To see specifically how the carbon price has impacted CO_2 emissions per fossil fuel, a distinction between coal, gas, and lignite will be done. This goes in line with the methodology followed by Gugler et al. (2021). To test whether the MSR has operated similar to the UK CPF, a model of Differences in Differences (DD) will be carried out using the same variables. This method was employed by Abrell, Kosch, and Rausch (2021).

Discussions about the introduction of a CPF to the EU ETS are on the table (Flachsland et al., 2020). Therefore, to determine whether its introduction makes sense on top of the MSR is the contribution of this research to the debate. The repercussions of an additional reform are enormous. Especially because the EU ETS is being followed by other countries. South Korea and the People's Republic of China (China) are two of them. South Korea released the South Korea Emissions Trading Scheme (KETS) in January 2019 to reduce its GHG emissions by 2030 (Winchester & Reilly, 2019). In China, the operations of its ETS started officially in 2021, after concluding a test phase (IEA, 2021). Both countries have followed the recommendations of the EU Commission such as the implementation of market stabilization policies. Still, only South Korea has stated its desire to implement a carbon price floor or ceiling in case of oversupply (International Carbon Action Partnership, 2021). The fact that South Korea may introduce a carbon price floor only in case of market oversupply, instead of introducing it as a permanent measure (like in the UK), validates the lack of consensus of its effectiveness. Since almost all countries aim to combat climate change, it is important to contribute to reply this open question.

The paper proceeds as follows: In the next section, the existing literature is presented. In the third section, the background about the EU ETS and the UK CPF as well as the electricity generation of each country are described. Then, the two hypotheses are presented. After that, the paper presents the data and the methodology in section five and six, respectively. In section seven, the empirical findings are discussed, and section eight concludes the study with the main findings and future research directions.

2. Literature review

Since the introduction of the EU ETS and the UK CPF, a rich body of literature reviewing the effectiveness of these policies to undertake GHG emissions has emerged. Specifically, mixed results about these two policies can be found in the literature. The results differ depending on the industries (Abrell, Faye, & Zachmann, 2011) and countries studied (Koch, Fuss, Grosjean, & Edenhofer, 2014), the time frame analyzed (Muûls, Colmer, Martin, & Wagner, 2016) and (Ellerman, Convery, & de Perthuis, 2010), and the methodologies used to determine the effectiveness (Declercq, Delarue, & D'haeseleer, 2011) and (Bel & Joseph, 2015). This document will present literature from the introduction of the EU ETS, in 2005, to the present year, 2021. However, studies from 2017 will be presented extensively because in that year starts the scope of this investigation.

In the following section, relevant studies about the EU ETS and the UK CPF will be presented. The literature is divided as follows: first, studies about the EU ETS in phases I (2005 - 2007) and II (2008 - 2012) are presented. Then, literature about the EU ETS in phase III (2013 - 2020) are introduced. Third, research that investigated the impact of the UK CPF in the British CO_2 emissions are discussed. Fourth, studies that compare the effectiveness of carbon pricing in the UK and Germany are presented. Finally, the research gap is explained.

2.1. EU ETS in phases I (2005 - 2007) and II (2008 - 2012)

During phases 1 and 2, events such as the over-allocation of EUAs and the economic recession have undermined the efficacy of the EU ETS (Oestreich & Tsiakas, 2015), (Laing, Sato, Grubb, & Comberti, 2013), (Abrell et al., 2011), and (Anderson & Di Maria, 2011). Accordingly, Declercq et al. (2011) and Bel and Joseph (2015) found that the reduction of GHG emissions during the recession (2008 - 2009) was caused by the economic crisis. It should be noted that both studies used different econometric methods. Declercq et al. (2011) used a counterfactual scenario that estimates how the fuel prices, electricity demand, and CO_2 price would have been affected if the economic recession had not happened. Then, the authors compared both scenarios. On the other hand, Bel and Joseph (2015) used historical emissions data as a baseline for their dynamic panel model. The indicators used in this analysis are a variable representing policies, the CO_2 emissions under the EU ETS, the electricity industry index, the price of electricity and gas, a dummy variable for the economic crisis, and the consumption of coal, natural gas, and electric energy. However, Abrell et al. (2011) agrees partially, concluding that the EU ETS impacted the reduction of GHG emissions. The authors analyzed the change in firms' emissions from the first to the second phase. The authors found that both changes in the economic activity and the changes in the EU ETS from the first to the second phase explained the reduction of GHG emissions. This suggests that the stricter rules imposed in phase 2 (2008 - 2012) as a lower cap, less free allocation, and higher penalties improved the effectiveness of the EU ETS. Hintermann, Peterson, and Rickels (2016) agree with this finding. Moreover, the authors add that the reduction of EUAs during the recession (2008 - 2009) shows that the instrument is flexible to adapt to market conditions while maintaining its value above zero. The authors reached these conclusions after analyzing the existent literature about the EU ETS, excluding studies about a carbon price floor. Nevertheless, the authors concluded that a CPF in the EU ETS would be less environmentally beneficial than reforms such as limited banking [1].

The studies discussed so far show that there is not a consensus about the effectiveness of the EU ETS to tackle GHG emissions during phases 1 and 2. On top of GHG emissions reductions, another key objective of the EU ETS is to promote clean investments. This was studied by Hoffmann (2007) and Rogge, Schneider, and Hoffmann (2011). After surveying agents of the power sector in Germany, both authors resolved that the EUAs were driving small - but insufficient investments in low-carbon technologies. In line with this objective, but opposed to what Hoffmann (2007) and Rogge et al. (2011) found out, the UK government determined in 2010 that the EU ETS alone was ineffective in reducing GHG emissions. Therefore, in December, the UK surveyed companies and individuals involved in the power sector to know their opinion on a carbon reform proposal (UK Government, 2010). This document addressed the need for a carbon price on top of the existing EUAs to promote long-term investments in low-carbon technologies. It noted that these technologies are essential to achieve the transition towards a greener future but are risky investments due to their higher risk and volatility compared with fossil fuels. After the consultation, the CPF was announced as an environmental tax in the Budget 2011 to become effective from April 2013 (UK Government, 2011).

2.2. EU ETS in phase III (2013 - 2020)

The policies of the EU ETS in phase 3 changed substantially. This phase introduced new sectors and aimed to increase the control of the new allowances (a detailed explanation will be found in section 3.1.3). At the same time, the UK CPF became effective in April 2013.

Discussions about the advantages that a price floor would represent for the EU were introduced by several authors. Koch et al. (2014) were among the first. The authors analyzed the period from January 2008 to October 2013 with an Ordinary Least Squares (OLS) model. The variables included were the price change of gas and coal, a theoretical switching price between gas and coal, the price change of the European stock exchange, and the electricity production growth from wind, solar, and water sources. They determined that the reformed EU ETS would be ineffective in promoting decarbonization because the EUA price was not significantly affected by demand shocks (e.g., economic recession). This finding challenged the results discussed before and undermined the effectiveness of phase 3 because its changes were focused on reducing these effects. Hence, the authors suggested setting a price floor, which will promote decarbonization by reducing the uncertainty of the dynamics of the EUA price. Accordingly, Edenhofer et al. (2017) supported this view and added that a price floor would reduce the regulation uncertainty, market myopia[2], and the waterbed effect[3].

Conversely, Gerlagh, Heijmans, and Rosendahl (2021) suggested that a further modification of the MSR could be good enough to improve its effectiveness. They analyzed the impacts of the MSR with a dynamic model of two periods. The variables included in the model were the supply of allowances, the interest rate, the elasticity of the emissions' demand, and parameters that estimate the banking effect,

[1] The banking policy allowed ETS participants to transfer their unused allowances from phase 2 to phase 3 (European Commission, 2015a).

[2] It is referred to the lack of long-term view by market participants. In the EU ETS, there is an absence of a minimum price that secures return over investments. Therefore, its design does not reduce market myopia, undermining investments in low-carbon technologies (Edenhofer et al., 2017) and (Schmidt, 2020). On the other hand, a CPF directly tackles this problem by securing a minimum price.

[3] It is when an opposite result is derived from an economic policy. In the EU ETS is caused mainly by two factors. First, because of its fixed cap. When companies reduce GHG emissions, the demand for EUAs decreases while the value of the cap is kept. Second, due to the MSR. The EUAs store in the MSR are expected to be bid later, instead of being eliminated. In both cases, the price of EUAs is negatively affected, undermining the effectiveness of the policy (Gugler et al., 2021), (Edenhofer et al., 2017), and (Schmidt, 2020).

and the cancellation policy. The authors proposed two revisions. On the one hand, that the MSR develops a hybrid price-quantity cancellation policy that cancels EUAs when the demand drops. At the same time, that the EUAs hold by MSR should be based on continuous rules rather than the discrete ones that are published yearly.

Another part of the literature focused on the impacts that the reformed phase 3 had on the endorsement of technologies that reduce GHG emissions. Eichhammer, Friedrichsen, Healy, and Schumacher (2018) studied these effects in the industries of cement clinker, pig iron, ammonia, and nitric acid, representing 40% of industrial emissions under the EU ETS. They had two interesting findings. On the one hand, they found that phase 3 had increased the incentives to adopt clean technologies. On the other hand, they found that by 2017, there was no evidence that the companies adopted these technologies, with nitric acid as the only exception. Finally, the authors stated that rising carbon prices - at that moment € 16 - will drive investments in low-carbon production processes. Perino and Willner (2016) looked into the impact of the MSR. The authors carried a dynamic optimization equilibrium model to study the MSR when it was proposed in 2015. Their approach took into consideration parameters such as banking, cost of abatement, allowances that declined at a constant rate, and an infinite time horizon. The authors concluded that the MSR is effective only when the markets perceive temporary scarcity - which is not always the case. About the low-carbon investments, they determined that its impact is ambiguous the EUA price is still uncertain. Both reasonings are compatible with the arguments exposed in the previous paragraph that support the establishment of a CPF - because it would promote long-term investments by securing a minimum carbon price.

Likewise, the Global Financial Crisis in phase 2, the Covid-19 crisis affects the EUA during phase 3. Gerlagh, Heijmans, and Rosendahl (2020) carried a study about the impact of this crisis on the MSR. After using a deterministic model to simulate an ETS market with and without the MSR, the authors concluded that the MSR is a good stabilizer. Nevertheless, the extent of it depends on the duration of demand shocks. The MSR works well for short-lived demand shocks, but not at all for long-lived demand shocks. By the end of their research, the type of shock that the Covid-19 crisis was, was not clear. The authors coincided that the dynamics that the MSR follows, are uncertain and that the introduction of a price floor would be a policy improvement.

2.3. UK Carbon Price Floor (2013 - 2020)

Another part of the literature focused on the effects of the CPF in the UK. Abrell et al. (2021) analyzed the impact of the UK CPF on the fossil-fired power plants from 2009 to 2016. The variables used by the authors were the hourly output by fossil fuel plants, fuel and carbon prices, the available hourly capacity, the residual demand, and the efficiency, emissions, and emission factor per power plant. They analyzed 35 plants of natural gas and 15 coal-fired plants. The authors used machine learning to predict the behavior of the power plants without the UK CPF. After creating the control group, they compared the GHG emissions with a Difference in Differences (DD) method. They found that, from 2013 to 2016, the UK CPF lowered the emissions by 6.2% at an average cost of € 18 per ton. One of the limitations of this paper is that it focuses only on short-term variables, excluding effects such as the investment in renewables and energy efficiency.

Likewise, Marion (2019) examines the same effect but considers the growth in wind and solar capacity, opted-out plants, and net imports of electricity. The author used the DD method to compare a synthetic UK power sector production per capita (created by weighting different European countries' production) with the real one. She tests the robustness of her estimation by running an "in-time" placebo and a permutation test. The author concluded that the UK CPF was a successful policy that reduced the GHG emissions of the power sector by a range from 41% to 49% over the 2013 - 2017 period. Also, she found that there was no increase in net imports. Both documents agreed that the UK CPF was significantly effective in reducing GHG emissions from the UK. Also, both papers recognized that the carbon tax was high enough that left many fossil-fired plants out of business. On the other hand, both papers lack of analyzing the impact of the UK CPF in driving low-carbon investments, which are fundamental to reach the zero target in GHG emissions of the UK. This paper will incorporate that analysis by measuring the impact of the carbon price in electricity generated by renewable sources.

2.4. Carbon Pricing in Germany compared to the UK

It is hard to compare the reduction of GHG emissions among different countries. Among the various reasons that emerged are the differences in energy sources, market interconnection, climate policies, and electricity price determination. Nevertheless, the UK and Germany have similar energy sources as well as the same electricity price determination (both will be discussed in section 3.4). Still, the countries have taken different climate policies in the last decade.

Gugler et al. (2021) compares the success of these countries' policies in encouraging the production of renewables. The authors examined the effects of the carbon price on CO_2 emissions from gas and coal, as well as on the production of wind and solar energy in the UK and Germany. The effects depend on the different climate policies that both countries exercised. The UK used a carbon pricing scheme, while Germany offered subsidies for renewables. First, the authors estimated daily CO_2 emissions from gas and coal plants after conducting two models: a Heckman two-step and an OLS. Then, they used the same model to derive its marginal effects on carbon pricing, and energy production from wind and solar. They got mixed results. On the one hand, they concluded that a carbon pricing scheme is more effective than renewable subsidies when its carbon price is high enough. For both countries, that means a carbon price above € 14/tCO_2. On the other hand, they concluded that these two policies together can be mutually enforcing in Germany but mutually

opposing in the UK. The reason behind this is that the carbon costs in Germany are low compared with the UK.

Similarly, Wilson and Staffell (2018) agrees that the carbon price encouraged fuel switching in the UK faster than in Germany. However, the method they followed is different. The authors analyzed the fuel-switching through data comparison. Nevertheless, both documents agreed that most of the British switch was towards gas and not renewables, which is the main objective. Still, they also recognized that a higher carbon price could replace gas for renewables. These findings support that the UK CPF has been more effective in reducing GHG emissions than the EU ETS alone because its price was higher.

2.5. Research gap

As explained in the previous paragraphs, several studies have examined the influence of the EU ETS and the UK CPF at different periods, industries, and among different countries. Also, most of the research has focused on one of these two policies, being scarce the studies that compared them. However, two factors are missing: 1) the consideration of the nuclear policy, and 2) the implementation of the MSR. The addition of the electricity generated by nuclear sources as an explanatory variable of the CO_2 emissions makes sense. This is sustained by the fact that only Germany had closed nuclear power plants since 2011. Also, because it is the only reliable carbon-free energy source that can generate electricity 24 hours a day, 7 days a week in a reliable way, as fossil fuels (Gates, 2021). If this factor is relevant, it will add value to the debate on the effectiveness of the UK CPF. Also, to test whether the introduction of the MSR could have acted as a CPF for the EU ETS is missing. The MSR operated as a good stabilizer during the Covid-19 crisis, where the EUA price maintained its value above € 16 and then quickly recovered despite the economic recession. To determine whether the EU ETS is as effective as the UK CPF due to the MSR is the second goal of this research. To do so, two models will be performed: an OLS panel data, and a Differences in Differences model.

Another difference is that this research compares the effectiveness of the EU ETS in Germany with the UK in the period from January 2017 to December 2020. A period when the EU Commission had implemented market stabilizer reforms, the UK CPF was in operation, Germany's nuclear phase-out was a reality, the MSR was announced (in 2017), and then implemented (in 2019).

The paper will continue as follows: In the next section, the background about the EU ETS and the UK CPF as well as the electricity generation of each country is described. Then, the two hypotheses are presented. After that, the paper presents the data and the methodology in section five and six, respectively. In section seven, the empirical findings are discussed, and section eight concludes the study with the main findings and future research directions.

3. Background on the EU ETS and UK CPF

3.1. Development of the EU ETS

Climate change is a problem that needs global cooperation to be effectively solved. For that purpose, the United Nations (UN) created the United Nations Framework Convention on Climate Change (UNFCCC) in 1994 (European Environment Agency, 2014). The UNFCCC organized and helped to monitor the 1997 Kyoto Protocol, the first globally legally binding agreement on GHG reduction that the EU ratified (European Commission, 2021a). The EU ETS was launched in 2005 to help the EU to meet its Kyoto targets, and later on, their 2015 Paris targets. The scheme is based on a cap-and-trade system, where the cap represents the GHG emissions that can be emitted by installations covered by the system. The trading principle allows the companies to trade EUAs within the cap. For emissions to decline, the cap is expected to decrease over time. At the end of each year, an installation must pay a penalty if it does not have enough EUAs to cover its emissions (Hirst, 2018). That means that if a company increases its production without decreasing its emissions, it must buy EUAs in the trading market. The participating countries and industries, the rate at which the cap decreases, and the penalty that participating companies must pay have changed throughout the different phases. All phases will be described in the incoming paragraphs. Nevertheless, an extended analysis will be carried out for phase 3. This is because the scope of this study is focused on the period January 2017 to December 2020 - which belongs to that phase.

3.1.1. Phase 1 (2005 - 2007)

The first phase of the EU ETS was a pilot phase where 27 countries participated. The penalty for non-compliance was set at € 40 p/ton. It covered the CO_2 emissions of power stations and other combustion plants ($\geq$ 20MW), oil refineries, coke ovens, iron and steel plants, cement clinker, glass, lime, bricks, ceramics, pulp, and paper and board (European Commission, 2015a). To avoid the risk that companies move their production abroad (carbon leakage), the EU issued almost 100% of the EUAs for free. This phase helped the EU to set a carbon price, to create infrastructure to monitor, report and verify the emissions, and allowed the free trade of EUAs (European Commission, 2021a).

The European Environment Commissioner Stavros Dimas concluded that the EU ETS was being successful because compliance rates were high and CO_2 emissions in 2006 increased by 0.3% below the economic growth, which grew by 3% (European Commission, 2007). By the end of the first phase of the EU ETS, it was not possible to clearly measure the impact on CO_2 emissions because of the lack of verified data (European Commission, 2007).

Nevertheless, this phase suffered from some difficulties. First, the EUAs were delivered based on wrong estimates - which later caused an oversupply. Second, the (almost) totally free allocation of the EUAs happened in an uneven way - which favored some firms over others. Abrell et al. (2011)

determined that non-metallic minerals were negatively affected in comparison to the other sectors. Third, many companies profited from the system without reducing CO_2 emissions. Sijm, Neuhoff, and Chen (2006) and Smale, Hartley, Hepburn, Ward, and Grubb (2006) demonstrated that power companies made windfall profits due to the EU ETS. At the end of the phase, the price of the EUAs was zero. Also, the EUAs not used could not be stored because banking was not allowed.

3.1.2. Phase 2 (2008 - 2012)

The second phase of the EU ETS was binding. It considered the targets of the first commitment of the Kyoto Protocol. Alternative ways of reducing emissions abroad were allowed through the Certified Emission Reductions (CERs)[4] and Emission Reduction Units (ERUs)[5] (European Commission, 2015a). Also, the phase added new features. First, three new countries participated: Norway, Iceland, and Liechtenstein. Second, the penalty for non-compliance was increased to € 100 p/ton. Third, the cap was reduced by 6.5%, the free allocation of EUAs fell to 90%, and their banking was allowed. Fourth, the aviation sector was included in 2012 - applying only to flights between airports located in the European Economic Area (EEA) (European Commission, 2021a). Finally, some countries took voluntary measures such as the inclusion of nitrous oxide (N2O) on top of CO_2 and auctioning.

As a result of the measures took to strengthen the EU ETS, the price of the EUAs increased during the firsts six months of 2008 until the Global Financial Crisis (GFC) hit. The recession (2008 - 2009) caused a contraction in global production that subsequently reduced the demand for EUAs. Figure 1 shows the falling of the EU ETS price from almost 30€ /tCO_2 in mid-2008 to less than 7€ /tCO_2 at the end of 2012. Despite the collapse of the price, the EU reduced its GHG emissions by 8% below 1990 levels. Thus, the EU exceeded the target of 5% (European Commission, 2021b). During this period, Germany and the UK reduced their emissions by 21% and 12.5%, respectively.

By the end of phase 2, there was an excess of two billion unused EUAs that could be banked to be used in phase 3. The EU considered the EU ETS as a good policy instrument that needed further reforms. Connie Hedegaard, European Commissioner for Climate Action stated that the EU ETS was reducing GHG emissions, but that the market oversupply was undermining its impacts on energy efficiency and green technologies (European Commission, 2012). In that same meeting, the EU Commission approved the delay of 900 million allowances that were supposed to be held in 2013. Conversely, the UK believed that the EU ETS reforms were not strong enough. After approving the CPF in 2011, they introduced it in April 2013 as an additional cost on top of the EU ETS to meet its goals towards decarbonisation (UK Government, 2011).

3.1.3. Phase 3 (2013 - 2020)

The third phase was also binding and summed up 31 countries after Croatia joined in 2013. It considered the targets of the 2nd commitment period of the Kyoto Protocol. The targets for 2020 were a 20% cut in GHG emissions from 1990 levels, a share of 20% in renewables, and an improvement of 20% in energy efficiency (European Commission, 2021c). This phase introduced many changes. First, the consignment of EUAs. The power industry was required to buy them via auctioning, while the industry and heating sectors received them for free (European Commission, 2015a). This occurred after the EU ETS Directive determined that companies of the power sector passed the cost of allowances to the consumers (European Commission, 2015a). Second, the cap started to decrease by 1.74% yearly. Third, the abatement solutions through the CERs and ERUs were reduced - meaning that domestic solutions were preferred (European Commission, 2015a). Fourth, the sectors of aluminum, petrochemicals, ammonia, nitric, adipic, and glyoxylic acid production, CO_2 capture, transport in pipelines, and geological storage of CO_2 were added (European Commission, 2015a). Finally, the inclusion of nitrous oxide (N2O) from all nitric, adipic, and glyoxylic acid production and PFC from aluminum production became mandatory (European Commission, 2015a).

Moreover, the EU ETS Directive made two relevant adjustments in this phase. The first was 'back-loading', a measure that postponed until 2019, the auction of 900 million of EUAs that were scheduled to be sold during the period 2014 - 2016 (European Commission, 2021a). This mandate allowed the reduction of the surplus of allowances generated after the GFC. The second was the Market Stability Reserve (MSR), drafted firstly in 2015, but confirmed in 2017, that operates from 2019 onwards (European Commission, 2017). The MSR allows the EU ETS Directive to control the volume of EUAs to be auctioned through a 'reserve and release' system (European Commission, 2021a). Initially, the MSR reserved the 900 million of EUAs from 'back-loading to then auction them. Subsequently, 12% of EUAs are reserved when the market has a surplus higher than 833 million. The MSR releases EUAs in yearly batches according to pre-defined rules that are published every year on May 15th (European Commission, 2021d).

In this period, the EU ETS Directive applied changes to strengthen the EUA price. Figure 2 shows eight events and the EUA price development during phase 3. Four of them are considered the most relevant. First, on 6th November 2013 when the all the participating countries ratified the second commitment period of the Kyoto Protocol (European Commission, 2013). The ratification confirmed the determination of the EU to comply with the climate international targets and to strengthen the EU ETS. Second, on 17th February 2015 when the Commission proposed to create the MSR and to become a world leader in the development and man-

[4] CERs are emissions certificates given by the UNFCC and the Kyoto Protocol after countries or companies successfully invest in sustainable projects in developing countries (European Commission, 2015a)

[5] ERUs are emissions credits granted to countries or companies after their complete Joint Implementation (JI) projects (European Commission, 2015a)

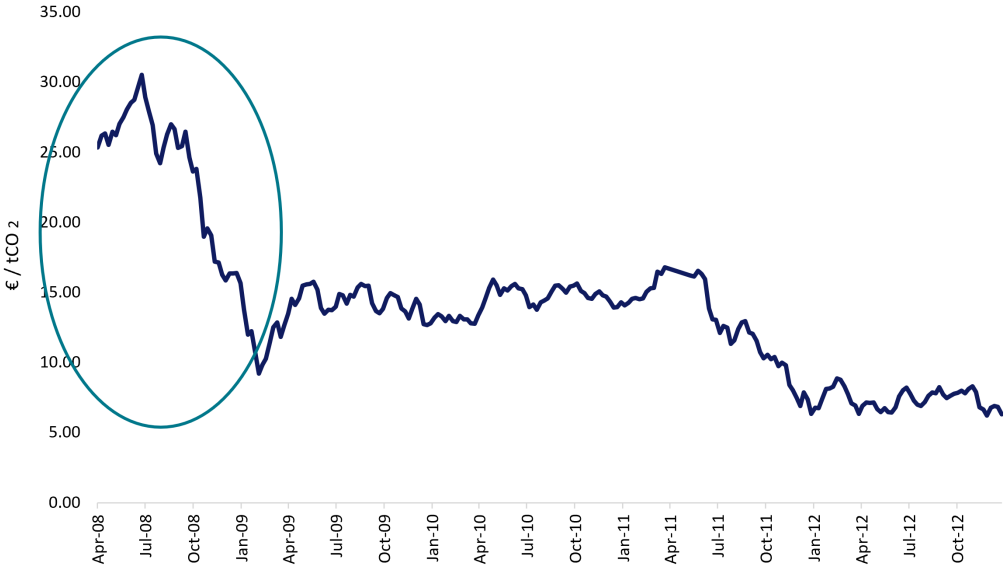


Figure 1: Weekly price development of EUA during phase II

Figure created by the author based on data provided by Sandbag (2021)

ufacture of renewable energy technologies (European Commission, 2015b). These two events were followed by a positive trend in the price. Third, on 9th November 2017, after two years of negotiations, the Commission approved the MSR, applied policies to tackle carbon leakage and support innovation and investment in clean technologies (European Commission, 2017). This policy supported the EUAs significantly, generating a continuity in the positive trend of the price. Fourth, on 17th September 2020, the Commission proposed to increase the reduction of GHG emissions 2030 target to at least 55% (European Commission, 2020). Finally, it is worth noting that the Covid-19 crisis affected the price only temporarily in contrast with the collapse generated during the GFC.

3.2. United Kingdom Carbon Price Floor

The UK introduced the United Kingdom Emissions Trading Scheme (UK ETS) in March 2002, three years before the EU (Bourn, 2004). The system was similar to the first EU ETS. Participating companies bid GHG emission reductions from 2002 to 2006 in exchange for a share of £215 million of national incentive funding (Bourn, 2004). Annual GHG emissions revisions were carried during the scheme. The companies' reduction target was calculated as an average of their GHG emissions from 1998 to 2000 (Bourn, 2004). In similarity to the EU ETS, some emissions were overestimated. Therefore, these companies could have received incentive payments even without reducing its GHG emissions. However, four participants that accounted for 50% of the incentive pool reduced their emissions considerably and stated that the system was effective. The UK ETS served many purposes. First, it established and created awareness of emissions trading in the market. Second, it secured 3.96 million tons of CO_2 emissions reduction. Finally, it influenced the design of the EU ETS (Bourn, 2004).

In 2009, was the first time that the introduction of a carbon price floor on top of the EU ETS was discussed in the UK (Marion, 2019). However, the Labour party opposed it. In 2010, the Coalition Government put it back on the table. Then, in December 2010, the UK government consulted companies and individuals of the power sector to get their opinion about a carbon pricing proposal (UK Government, 2010). The consultation made some remarks. The unstable and not high enough price of the EUAs had weakened investments in low-carbon technologies (UK Government, 2010) & (Marion, 2019). Renewable energy was more expensive and had higher exposure to price volatility than fossil fuels. Still, substantial investments were required in renewables, carbon capture and storage (CCS), and others to meet their sustainable goals. The Government's objective was to reduce 236 MtCO_2 over all sectors between the periods of 2008-2012 and 2013-2017 (Marion, 2019). Regarding a carbon price, the proposal was to combine the existing EU ETS plus price support. Specifically, it outlined three combined carbon prices (EUA plus UK CPF) of £20, £30, and £40/tCO2 in 2020 that will increase in 2030 to £70/tCO2 (UK Government, 2010). These estimations were based on a carbon price that will keep the increase of global temperature below 2°C. In the Budget of 2011, the Government approved a Carbon Price Support (CPS; also known as CPF) for electricity generation of £16/tCO2 that will reach £30/tCO2 in 2030 (UK Government, 2011). The policy started in April 2013. The tax rate per tCO2 was applied in addition to the EUA price and was expected to increase yearly. This rate will depend on the estimated EUA price (Marion, 2019). In the end, the CPS discontinued its increment after the period 2015-2016 because

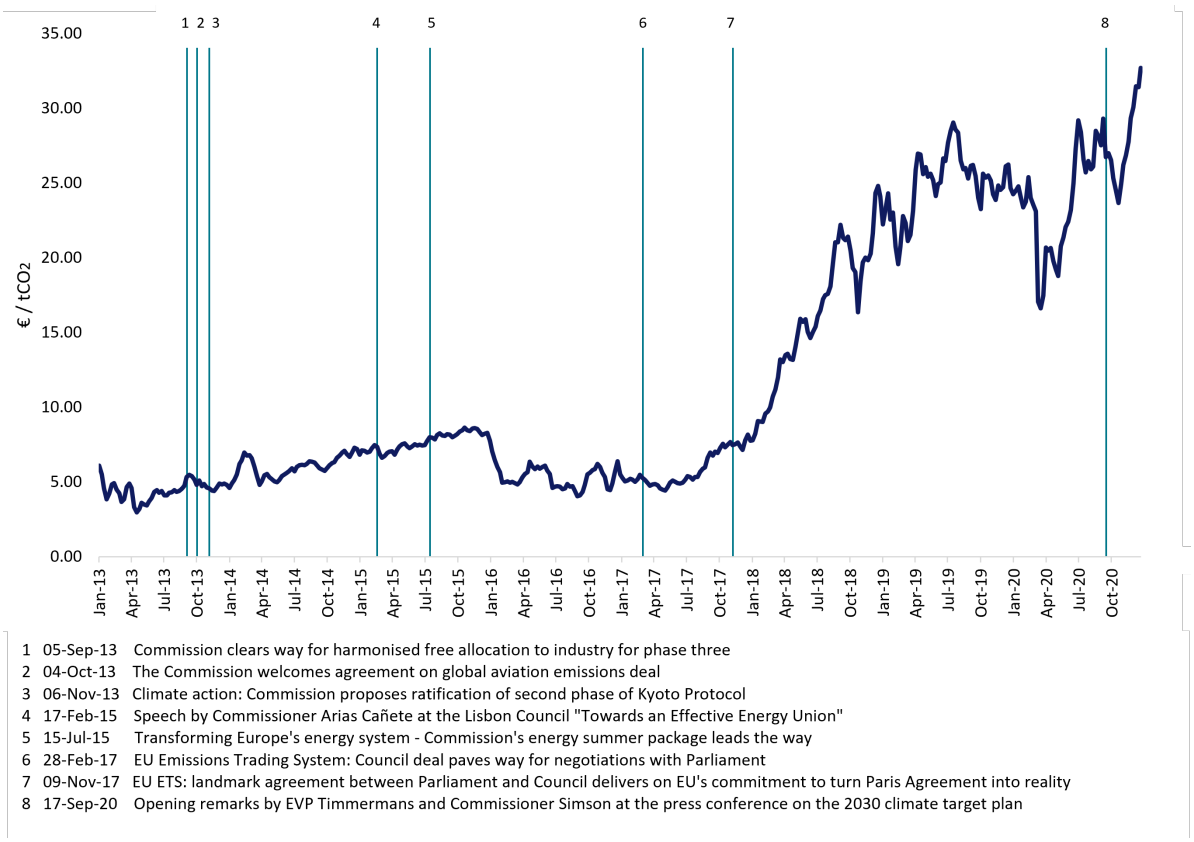


Figure 2: Weekly price development of EUA and events during EU ETS phase III

Figure created by the author based on data provided by Sandbag (2021) and the EU Commission (2021)

business representatives complained about their competitive loss. Both industrial and consumers pay higher rates for electricity than other European members. As a result, in May 2014, the European Commission approved compensation for some British electricity users for the extra costs produced by the CPF (European Commission, 2014). The EU Commission agreed that the CPF policy was in line with the goals set in the Environmental and Energy Aid Guidelines, and it was not distorting the competition with the block. As mentioned before, the CPS rate did not increase as announced. Table 1 shows the CPS freeze carried by the UK government in 2015. Because the CPS rate is based on the carbon content of the fuel used for power generation, coal plants were the most affected ones. The system also included Combined Heat and Power (CHP) operators and auto-generators (Marion, 2019). In general, all generators with a thermal input higher than 2 MWth had to pay the CPS. (Marion, 2019) calculates that on average the CPS rate on coal plants is 70% higher in comparison with the tax on natural gas. Likewise, the UK Government estimated that the impacts on Energy Intensive Industries (EIIs) such as steel and chemicals oscillated between 1% and 50% depending on their dependence on fossil fuels (Hirst, 2018). The UK government assures that the UK CPF had decreased the coal production and encouraged the closure of many coal plants. These results will be discussed in the empirical results section of this document.

Table 1: UK CPS rates

Date	CPS Rate per tCO_2
2013-2014	4.94£
2014-2015	9.55£
2015-2016	18.08£
2016-2020	18£

Source: Hirst, 2018

3.3. Electricity production per source

The power generation mix refers to the generation of electricity by different energy sources. It excludes the energy used for transportation and large divisions of housing and industry. Globally, electricity generates 27% of GHG emissions; heating, cooling, and refrigeration 7%; agriculture, and livestock farming 19%, transportation 16%; and cement, steel, and plastic factories together 31% (Gates, 2021). Why is the electricity mix the focus of this study? Because the decarbonization of electricity is the most important one to meet the environmental global goals. Clean electricity can replace

the electricity generated from fossil fuels to transport people, cool buildings, and produce products (Gates, 2021).

However, one of the largest challenges is that clean electricity needs to be generated reliably. That means, as long as large-scale storage is not available, electricity generation must not depend on weather conditions or time of the day. Renewable sources such as hydro, wind, solar, and biomass cannot ensure that right now. Gates (2021) states that nuclear is the only carbon-free[6] source that can produce electricity 24 hours a day, 7 days a week, and that can be installed everywhere. Because of its importance as a carbon-free energy source, nuclear is included in this study. However, the political view on this technology is the main difference between the countries analyzed. Therefore, this difference and its implications will be explained in the incoming paragraphs.

Throughout the years, both countries have increased the share of renewables to more than 40%. However, the way both countries have achieved it, and the incoming challenges to meet their environmental goals are different. These would be explained in the following paragraphs. The analysis is divided among the three phases of the EU ETS. Therefore, years from 2005-2020 are taken into consideration, a longer period than the analysis of this study. However, the periods analyzed will show that most changes had happened in the third phase. In Germany, especially, a massive replacement of fossil fuels materialized between 2017-2020, which coincides with the scope of this study. Also, the changes in the CO_2 emissions per capita will be presented. The division of the CO_2 emissions per person allows the comparison of Germany and the UK in a comprehensible way.

3.3.1. Germany

In 2005, when the EU ETS was implemented, Germany had few renewables on their energy mix to generate electricity. Wind, solar, hydro, and other renewables represented 10.3%, while nuclear energy, 26.4%. The relationship between renewables and nuclear energy has reverted throughout time. Figure 3 shows this development. By 2020, the share of renewables in electricity generation had quadrupled to 44.9%, while nuclear had decreased to 11.3%. From 2005 to 2019, CO_2 emissions per capita of Germany decreased by 20.84% (Our World in Data, 2021).

The main reason for the decrease in nuclear production is that its phase-out became a reality in 2011 after the Fukushima Disaster (Clean Energy Wire, 2021). The government shut down eight nuclear reactors and approved to cease the rest of them by 2022 (World Nuclear Association, 2021a). According to the World Nuclear Association, by March 2021, Germany had 6 reactors in operation and has closed 30. The gap left by nuclear energy is expected to be met with natural gas production and imports (IEA, 2021). The latter adds pressure on Germany to meet its environmental goals, which include being carbon-free by 2050. Another reason that explains the shift is the promotion of renewables to substitute both coal and nuclear energy that is part of Germany's energy transformation (*Energiewende* in German). As a consequence, the German Government has subsidized investments in renewable energy (Gugler et al., 2021). For instance, Germany offered low-interest loans to anyone interested in installing solar panels and paid a feed-in-tariff (a fixed price) to anyone who generated it in excess (Gates, 2021). Also, the EU ETS is included as part of the *Energiewende* as an important policy.

At the end of phase I (2005 - 2007), electricity generation increased by 3%. In the generation mix, the share of nuclear energy decreased by 4.3%, wind and other renewables increased by 1.9% and 1.5%, respectively. No significant change was registered for coal nor oil, while gas increased by 0.5%. CO_2 emissions per capita of Germany decreased by 1.74% during that time (Our World in Data, 2021). All these minor changes in electricity generation mix and CO_2 emissions happened during a stable policy period.

During phase II (2008 - 2012), electricity generation decreased by 2%. In the generation mix, the share of solar increased significantly. It ended up representing 4.2%, after increasing by 3.5%. Consequently, the average cost of photovoltaic rooftop systems decreased by 62.5%, from 4000 € /kWp in 2008 to 1500 € /kWp in 2012 approximately (Wirth, 2021). The share of wind energy grew as in the previous period by 1.8%. The share of nuclear energy fell by 7.4% after the closure of six reactors. Finally, the share of coal increased by 1.2%, while gas decreased by 1.7%. CO_2 emissions per capita of Germany decreased by 4.67% during that time (Our World in Data, 2021).

Phase III (2013 - 2020) experienced most of the transformation. The electricity generation decreased by 10%, mainly because of gains in energy efficiency. In the generation mix, the share of renewable energy increased by 20.9%, with wind growing by 15.4% and solar by 4.1%. Wind energy experienced changes in different directions. It grew by 5,000 MW in 2017, but only by 280 MW in the first half of 2019 (Deutsche Welle, 2019). The slowdown is a consequence of wind's decreasing popularity among the citizens who live around the wind farms. New projects' permits have become slower to get due to new rules and longer approval times, which have increased from six months to more than two years (Deutsche Welle, 2019). That is a challenge for the *Energiewende* because wind energy is supposed to represent 65% of the energy mix. On the side of fossil fuels, the share of coal decreased by 21.8% after both lignite and hard coal have decreased significantly in 2019. This reduction is a consequence of less production rather than plant closures (Carbon Brief, 2019). Consequently, the EUA price almost quadrupled from 2017 to 2019. CO_2 emissions per capita of Germany decreased by 17.93% during that time (Our World in Data, 2021).

3.3.2. United Kingdom

In 2005, fossil fuels (coal, gas, and oil) generated three-quarters of the electricity of the UK. The share of nuclear en-

[6]Not all authors refer to nuclear as carbon-free, but in this paper, we take Gates (2021) approach. His approach considers that nuclear energy needs uranium as a fuel, which is a carbon-free source.

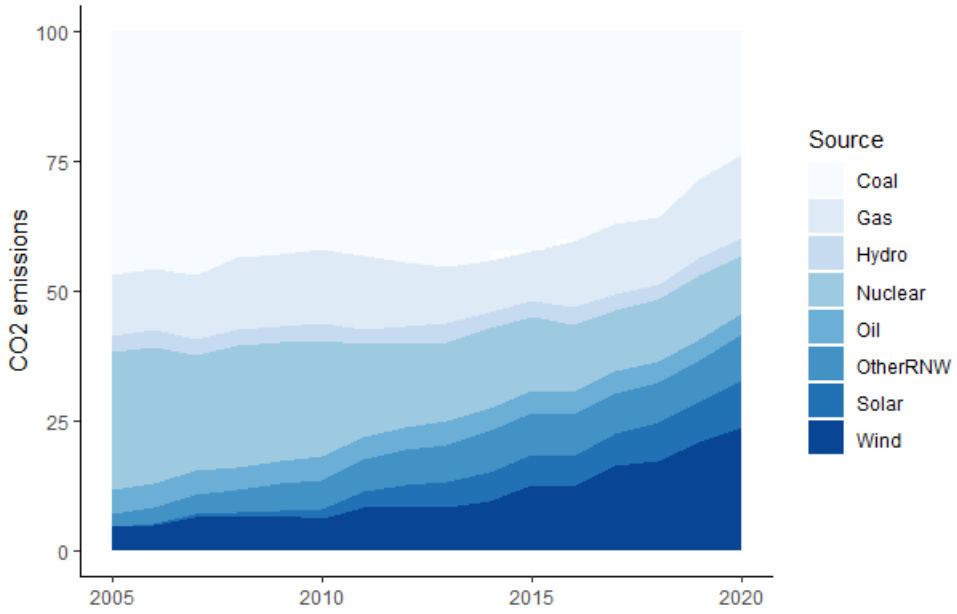


Figure 3: Germany: electricity production by source

Source: Our World in Data based on BP Statistical Review of World Energy & Ember
Figure created by the author

ergy was 20.6% and renewables less than 2%. In the same way as Germany, this relationship has changed. Figure 4 shows this development throughout time. Figure 4 shows that since the UK CPF introduction in 2013, the share of coal energy has been undertaken mainly by gas, solar, and wind energy. By 2020, renewables generated 42.4% of the electricity in the UK. However, nuclear energy produced 17% of the electricity in the same year. In contrast to Germany, the British government supported nuclear energy and considers it an important source to meet its climate goals. Currently, the UK is building two nuclear reactors and has strengthened measures to provide long-term support to investors (World Nuclear Association, 2021b). In the same way, natural gas has kept its share of electricity generation, and it is actively supported by the British government. Wind energy grew significantly, increasing from less than one percent in 2005 to 24.2% in 2020. From 2005 to 2019, CO_2 emissions per capita of the UK decreased by 42.06% (Our World in Data, 2021).

During phase I (2005 - 2007) electricity generation decreased by one percent. In the generation mix, the share of gas increased by 3.6%, it ended up representing 42.2% of the total. As shown in figure 4, gas undertake the electricity generated by coal. On the other side, the share of nuclear energy decreased by 4.6% after its generation changed from 81 TWh to 63 TWh. This happened during a positive context when the British government approved supportive measures for the industry in 2006 (World Nuclear Association, 2021b). CO_2 emissions per capita of the UK decreased by 3.74% during that time (Our World in Data, 2021). This exceeded Germany's reduction by 2%.

During phase II (2008 - 2012), the trend that favored natural energy against coal changed. At the end of this phase, the share of coal was 39.6% after increasing by 7.3%, while the share of natural gas decreased by 18%. Figure 4 shows how coal replaced gas from 2012 to 2014, the time that this shift lasted. It took over the higher share that gas earned from 2005 to 2010. The turning point was the suspension of fracking for several months in 2011 after it was proved that the method caused low-intensity earthquakes in Lancashire (BBC, 2012). The affected company resumed its operations in December of 2012 after the British government established additional preventive measures. On the other side, the share of nuclear and wind energy increased by 5.9% and 3.6%, respectively. Certainly, these policies impacted CO_2 emissions per capita of the UK, which decreased by 13.77% (Our World in Data, 2021). British reduction was approximately the triple of the one experienced by Germany. In the same way as Germany, the UK experienced many changes in phase III (2013 - 2020). Electricity generation decreased 14% after gains in energy efficiency. The UK is one of the IEA's leading countries in energy efficiency per GDP due to its policies in the modernization of buildings, transportation, digitalization, and others (IEA, 2019). By 2020 the share of coal on the generation mix was 1.7% after decreasing from 36.7%. Regulations imposed by the Government such as the UK CPF made coal an unprofitable industry since 2015 (IEA, 2019). The gap left by coal was covered by wind, gas, and other renewables after their share of generation grew by 16.2%, 9.6%, and 6.5%, respectively. However, in the long term, the UK expects to reduce its dependency on gas and increase

the production of renewable energy (IEA, 2019). Electricity generated by nuclear decreased by 27% because of plant closures. The Government considers nuclear as fundamental for the country, and the technology will increase its development in the long term. CO_2 emissions per capita of the UK decreased by 25.52% during that time (Our World in Data, 2021).

3.4. Carbon Price Comparison (2013 - 2020)

Since the introduction of the CPF, the price that the British had to pay for electricity increased substantially. Figure 5 shows this difference in Euro per ton of CO_2. In 2013, British consumers and companies pay double per ton of CO_2 than their European counterparts. This relationship oscillated throughout time. In 2016, the UK CPF was equivalent to 4.12x of the EU ETS. However, the final price per electricity did not increase in these rates because the British electricity generation reduced its dependence on coal, the most CO_2 intensive energy source. From 2013 to 2015, the share of coal in British electricity generation decreased by 14%, from 36.6% to 22.6%. At the same time, electricity generated by solar, wind, and other renewables increased by 9.1%. In Germany, electricity generated by coal decreased only by 3%. However, in the UK, the largest reduction in coal production happened after April 2015, when the Government duplicated the carbon price support. Figure 5 shows that, in average, the CO_2 price in the UK was 30 € /tCO_2, while in Germany, it stayed below 10 € /tCO_2. As a consequence, many British coal-fired plants closed. From 2013 to 2020, the generation of electricity from coal in the UK decreased by 35% from 36.7% to 1.7%. In Germany, where the price of EU ETS also increased, but less compared to the UK CPF, the coal generation decreased by 21.9% from 45.5% to 23.7%. These numbers show that the effectiveness of the CPF policy, which directly increases the marginal cost of fossil-fired power plants, is high (Abrell et al., 2021) and (Marion, 2019).

The paper will continue as follows: In the next section, the two hypotheses are presented. Then, the paper presents the data. In the section six the methodology is described. In section seven, the empirical findings are discussed, and section eight concludes the study with the main findings and future research directions.

4. Hypotheses

4.1. H1: There is a larger and significant reduction of GHG emissions due to the UK Carbon Price Floor than only with the EU Emissions Trading Scheme.

Some authors have investigated the impacts of carbon pricing in tackling GHG emissions in Germany and the UK. The comparison of these two countries is well-founded since both had similar electricity generation mixes before the introduction in the UK of the Carbon Price Floor in 2013. The electricity mix of both countries has changed. Currently, the dependency of the UK on coal for its electricity generation had decreased substantially. In 2020, there were 11 days where coal did not generate electricity in the UK. In that same year, coal generated only 1.7% and 23.7% of electricity in the UK and Germany, respectively. This happened while the price of carbon in the UK has significantly higher than in Germany. Flachsland et al. (2020) stated that the EU should establish a price floor for the EUAs because it 1) will increase its effectiveness as a policy tool and 2) it will provide credibility to green investments. Early on, in 2010, the UK Government agreed on both points. It declared that a carbon price floor is fundamental to promote long-term investments in low-carbon technologies (UK Government, 2010). Still, the share of electricity generated by renewables is similar in the UK and Germany. Renewable subsidies given by the German Government helped to close the gap left by a low carbon price (Gugler et al., 2021).

Nevertheless, the CO_2 emissions are lower in the UK than in Germany. Two factors explain this outcome. First, the UK switched from coal to gas, which emits less CO_2 emissions. According to Gugler et al. (2021), Abrell et al. (2021), Marion (2019), and Wilson and Staffell (2018); the UK CPF was effective in tackling CO_2 emissions. Second, the UK did not phase out nuclear energy, a process that Germany started in 2011. This second point has not been investigated by the existent literature. In this paper, the electricity generated by nuclear will be added as an exogenous variable, because it is also carbon-free. The decrease in electricity generated by nuclear may be a relevant factor that explains German CO_2 emissions. Especially because nuclear energy has been replaced by other fossil fuels (IEA, 2021). If the coefficient in the model is negative and significant, it means that less nuclear energy increases the CO_2 emissions of Germany. This effect, which is independent of the carbon price, would give an alternative explanation. It would mean that even if Germany had a carbon price floor, its effectiveness could have been undermined by the nuclear phase-out policy.

Table 2 presents the correlations between daily variations of the CO_2 emissions per fossil fuel and the nuclear electricity production per country. Also, it shows the correlations between the carbon price per country and the CO_2 emissions per fossil fuel. The time frame used is from January 2017 to December 2020. The chart shows that, in Germany, there is a high and positive correlation between electricity generated from nuclear and CO_2 emissions from coal, gas, and lignite. In the UK, it shows a positive but mild relationship. Hence, the chart justifies the inclusion of nuclear electricity as a relevant exogenous variable to explain the CO_2 emissions, especially in Germany. However, the positive correlation does not support the view that nuclear phase-out may have undermined the effectiveness of the EU ETS. Finally, the correlations show in the Table 2 show that the carbon price has a positive relationship with CO_2 emissions in Germany, but not in the UK. When we see this information in isolation, we can conclude that the carbon price has been effective in tackling the CO_2 emissions only in the UK. This supports the view of Flachsland et al. (2020). However, since there are other factors (i.e. electricity generated by other renewables,

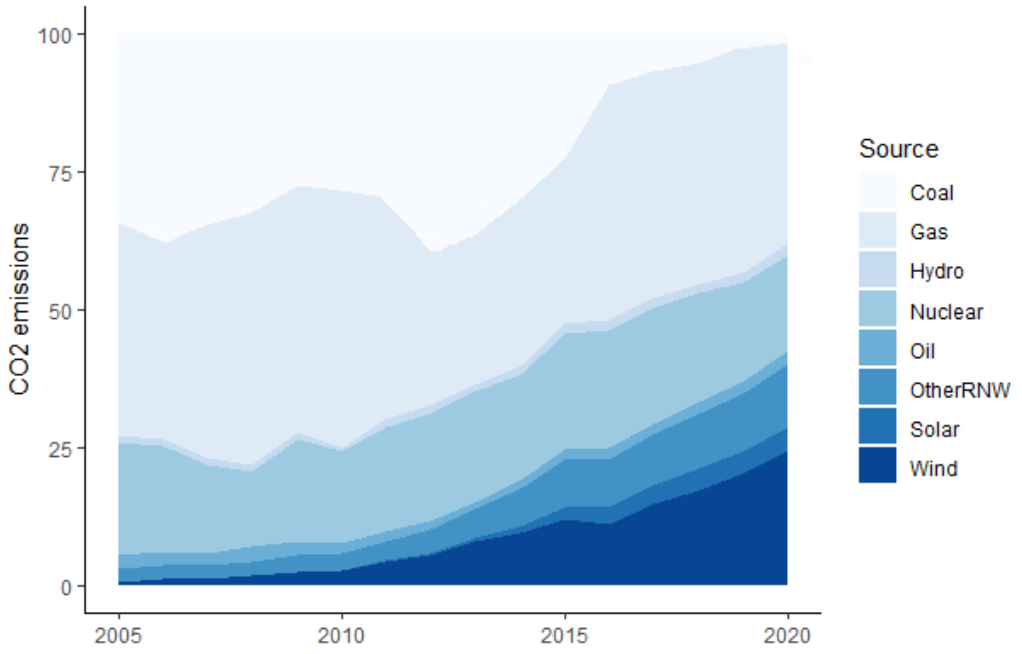


Figure 4: United Kingdom: electricity production by source

Source: Our World in Data based on BP Statistical Review of World Energy & Ember
Figure created by the author

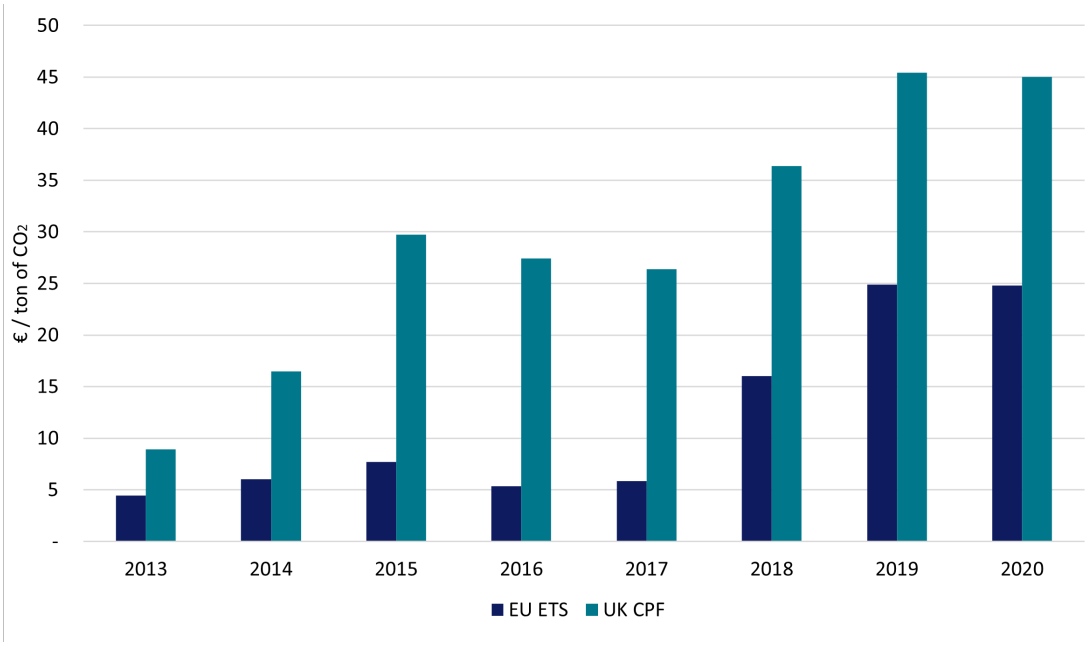


Figure 5: Yearly average of the EUA price and UK CPF

Figure created by the author

Table 2: Correlations between CO_2 emissions and other variables

	Germany	The UK
CO_2 emissions of Gas and Nuclear electricity	34%	1%
CO_2 emissions of Coal and Nuclear electricity	38%	6%
CO_2 emissions of Lignite and Nuclear electricity	54%	NA
CO_2 emissions of Gas and Carbon Price	1%	-6%
CO_2 emissions of Coal and Carbon Price	3%	-1%
CO_2 emissions of Lignite and Carbon Price	2%	NA

Coal-to-Gas ratio, and others) that affect CO_2 emissions, an econometric analysis that includes all of them is necessary to make relevant conclusions.

4.2. H2: The EU Emissions Trading Scheme is more effective since the Market Stability Reserve implementation

During phase 2 (2008-2012), the EU ETS lost credibility. The 2008 GFC caused a price collapse, and the instrument lost almost 70% of its value. Many studies analyzed the effectiveness of the EU ETS during this crisis. Declercq et al. (2011) and Bel and Joseph (2015) determined that the EU ETS was not resistant to economic shocks. On the other hand, Abrell et al. (2011) concluded that the EU ETS was still slightly effective in periods of economic crisis. After that, the EU Commission implemented new rules that made the EU ETS more resilient. The one of interest in this study is the MSR, introduced in 2019.

As mentioned in section 3.1.3, the MSR allows controlling the volume of EUAs that are in the market (European Commission, 2021a). Therefore, it is designed to avoid oversupply, but its future path is still uncertain because the MSR reacts to the market. Thus, some authors believe that the introduction of a CPF would be more effective to reduce uncertainty and promote long-term investments in clean technologies (Flachsland et al., 2020).

Nevertheless, the price of the EUAs during the Covid-19 crisis was resilient. In contrast with the 2008 GFC, where the price fell from 30 € /tCO_2 to 7 € /tCO_2 and did not recover; during the Covid-19 crisis, the EUA price fell from 26 € /tCO_2 to 16 € /tCO_2, but regained its previous value after four months. Moreover, the EUA price continued its uptrend and market a new high in July 2020, sustained by the support of the EU to its 2030 climate goals. Figure 5 shows that during 2015 and 2017, where most coal power plants were closed, the average CO_2 price in the UK was between 25 € /tCO_2 and 30 € /tCO_2. In Germany, during 2019 and 2020, when the MSR was in operation, the average CO_2 price was between 25 € /tCO_2. Could this policy have acted as a CPF for Germany? Since this discussion is on the table, therefore, an analysis of whether the MSR had acted as a CPF is important for research purposes.

Figure 6 shows the development of the average CO_2 of Germany and the UK and highlights the time during the MSR. Also, it shows four simple linear regressions, two for Germany and two for the UK in periods before and after the MSR implementation. This figure displays that the slope of the average CO_2 emissions of Germany changed after the MSR introduction. Even though the UK also continued to reduce CO_2 emissions, a pronounced shift of slope can be seen only in Germany. Before the MSR, in Germany, the trend of CO_2 emissions was slightly positive (+0.11). In the UK, it was neutral (+0.01). After the MSR, the value of these relationships changed. That period is shaded in grey. The slope of Germany's CO_2 emissions evolved to -0.35, i.e. it reduced by a factor of 4x. On the other hand, in the UK the slope only changed to -0.09. Nevertheless, factors such as the development of renewable energy, the coal-to-gas price, carbon price, and economic growth also influence the development of CO_2 emissions in each country. Therefore, an econometric model that includes these factors is needed to validate this hypothesis. For that purpose, the model of Differences in Differences will be performed.

The paper will continue as follows: In the next section, the data is described. In the section six, the methodology is described. In section seven, the empirical findings are discussed, and section eight concludes the study with the main findings and future research directions.

5. Data

This analysis covers the phase III of the EU ETS from 3 January 2017 to 31 December 2020. This period captures different economic developments and policy reforms. In 2017, the price of the EU ETS oscillated between 4.43 - 8.16 € /ton. It increased by € 2.61, in line with the uptrend of the STOXX 600, Europe's market index, which increased by 8.5%. In November of the same year, the EU Commission agreed to strengthen the EU ETS to fulfill the Paris Agreement (European Commission, 2017). As a consequence, the EUA price started an uptrend, which continued due to the introduction of the MSR in 2019. On the other hand, the UK kept its carbon price support at £18.08 during the whole period. Finally, the economic crisis due to Covid-19 started in February 2020, is also captured by the period analyzed.

All the prices used are expressed in Euros. The data used differs depending on the model. For the OLS in panel data, it consists of 1458 observations, which is a robust number for the econometric analysis performed. Table 4 shows the returns of the variables, which are used in the model because the panel data OLS needs stationary variables to be

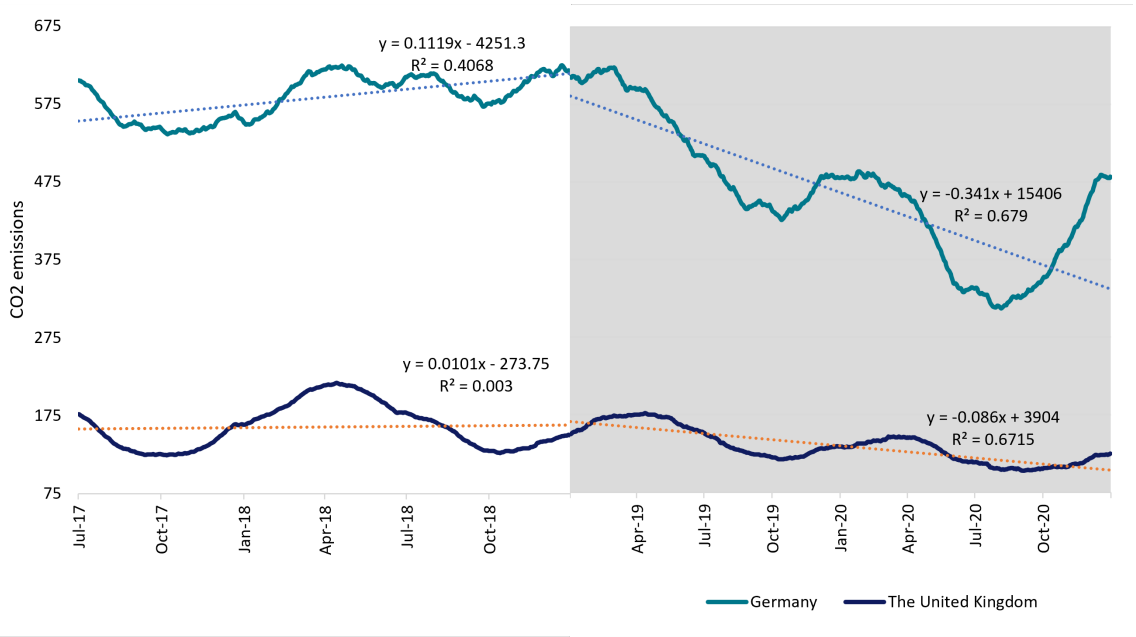


Figure 6: Semi-annual average of CO_2 emissions of Germany and the UK before and after the Market Stability Reserve

Figure created by the author

performed correctly. For the DD model, the output is expressed in differences, therefore level data is needed. The data used is shown in table 3, it consist of 1459 observations. In general, German variables are more volatile than the British ones. Except for the CO_2 emissions of coal in the UK, which have oscillated between 0 and 219 thousand tons, since there are 96 days where coal did not generate electricity in the UK.

Electricity generated by nuclear sources differs between the two countries. Despite the UK produces approximately 50% less electricity than Germany, its nuclear production is as large as the one of Germany, and the Government plans to promote it in the future. The electricity generated by solar and wind sources diverges as well. On average, Germany duplicates the share of the UK in electricity generated by both solar and wind onshore. In summary, nuclear energy in the UK represents the double than in Germany. However, Germany produces twice electricity from solar and wind onshore than the UK. For the purpose of the second analysis with the DD model, the electricity production is grouped by CO_2 *neutral sources*. It includes electricity produced by wind onshore, wind offshore, solar, and nuclear.

EUA price: The EUA price used is emitted by the ICE Futures Europe ECX. It is a continuous contract based on spot-month calculations. In this contract, each participant must either make or take delivery of the EUAs at the expiration date (Dhamija, Yadav, & Jain, 2018). As futures trade in higher volumes than spot carbon emissions they are more liquid (Dhamija et al., 2018). The EUA price is obtained from Sandbag, a non-profit think tank that focuses on climate change. Authors such as Abrell et al. (2021) and Marion (2019) have used the EUA future prices, as well.

Electricity demand by source: The electricity demand is obtained from the European Network of Transmissions System Operators for Electricity (ENTSO-E). The quantities used are under the denomination of Actual Generation per Production Type. They were found in 15-minute frequency for Germany and 30-minute frequency for the UK. Both were expressed in gigawatts and transformed to gigawatts per hour (GWh). The data that is divided per production type allowed the differentiation of two more variables: 1) the electricity generated by nuclear, solar, wind offshore and wind onshore; and 2) the CO_2 emissions. The CO_2 emissions were calculated using the CO_2 emission factors provided by *Umwelt Bundesamt*. These factors have a yearly frequency and are differentiated by fossil fuels: coal, gas, and lignite.

Stock Market Prices: National stock market indices are used as economic variables. For Germany, the DAX 30 is used. This index represents the 30 largest companies listed in the Frankfurt Stock Exchange. It is traded in euros and has high liquidity. For the UK, the FTSE 100 is employed. It represents the 100 largest companies listed on the London Stock Exchange. It is traded in British pounds, but in this document is valued is converted to euros. Both indices are total return indices. Thus, they include dividends. The data of both indices and the exchange rate are obtained from Investing in a daily frequency.

Coal-to-Gas Price Ratio: Finally, the Coal-to-Gas price ratio is included in the analysis. The coal price used is the Rotterdam Coal Futures (ATW). Each contract represents 1,000 metric tons of thermal coal. It is expressed in US Dollars. The natural gas price used is the UK Natural Gas Futures (NBP).

The contract size is 1,000 therms of natural gas, which are equivalent to 29,307-kilowatt-hours. It is expressed in British pounds. Both prices are first converted to Euros and then transformed to Euro/MWh. Since the final calculation is a ratio, the units are 1. All prices and exchange rates are obtained from Investing.

The paper will continue as follows: In the next section, the methodology is described. In section seven, the empirical findings are discussed, and section eight concludes the study with the main findings and future research directions.

6. Methodology

This paper has two main objectives: first, to show whether the UK CPF has been more effective in tackling the CO_2 emissions in the UK than the EU ETS in Germany. A coefficient between the carbon price and the CO_2 emissions will be calculated to determine the magnitude of this relationship in each country. Second, test whether the effectiveness of the EU ETS has increased since the implementation of the MSR in 2019. For this purpose, the difference between the CO_2 emissions of Germany and the UK will be estimated with a model of Differences in Differences (DD). The DD method is a variation of the linear panel data. This model will evaluate the change of GHG emissions in these two countries since the MSR was introduced. For this purpose, a dummy variable will be created. The MSR is a reform of the EU ETS to reduce oversupplies and to make the instrument resilient to economic shocks. Therefore, the consideration of the Covid-19 crisis helps to prove the last point. Finally, the day of the week effect is being considered for both countries.

Daily returns of energy and economic variables are used. The daily returns are calculated as (i) $r_{i,t} = ln(P_{i,t}) - ln(P_{i,t-1})$, where $P_{i,t}$ is the price of the index i at time t. This approach goes in line with (Gugler et al., 2021), who test the effectiveness of carbon pricing in Germany and the UK. The authors estimated the CO_2 emissions of Coal and Natural Gas power plants. Still, the difference is that in this paper, daily returns are used. In this paper, the CO_2 emissions are calculated based on the national electricity generation. For that purpose, the energy variables employed are returns of natural gas, coal, EU ETS and UK CPF, the electricity demand, the electricity production from solar, wind, and nuclear sources; and CO_2 emissions of Coal, Natural Gas, and Lignite. The Coal-to-Gas price ratio has been used by (Gugler et al., 2021), (Abrell et al., 2021), and many others because it represents the cost relationship between the two most important electricity fuels. (Gugler et al., 2021) and (Koch et al., 2014) utilized the production from renewable sources in their models, as well. The electricity production from renewables is relevant because their marginal cost is (almost) always lower compared to the one from fossil-fuel power plants. Therefore, they are ranked first in the merit order curve. The economic variables used are the prices of the Financial Times Stock Exchange 100 (FTSE 100), which represents the 100 biggest companies listed in the London Stock Exchange, and the *Deutscher Aktien Index* (DAX), which represents the 30 largest companies listed in the Frankfurt Stock Exchange. Several authors have included economic variables in their analysis of the carbon price. For example, (Koch et al., 2014) employed the returns of the European stock exchange and concluded that the EU ETS was not affected by demand shocks. Still, there is not a homogeneous consensus of the effects of an economic recession on the carbon price.

As the CO_2 emissions are time varying and are caused by many factors, it is crucial for its correct modeling to 1) identify the variables that influence on them and 2) use enough data that allows the application of the central limit theory. For both points, the two models of linear regression for panel data used in this paper are useful (Phillips & Moon, 1999). For the first point, both models applied in this paper will use five exogenous variables to estimate the CO_2 emissions of each country. The results of the models are coefficients that show long-run average relationships between the variables tested (Phillips & Moon, 1999). Since the both hypotheses of the model are to test whether the Carbon Pricing of the UK has been more effective in tackling the CO_2 emissions of the UK in comparison with the one implemented in Germany, the linear regression for panel data answers precisely that. For the second point, the data use in this paper are daily returns and daily values that make up a total of 1458 and 1459 observations, respectively, for each of the models and countries. The advantage of the linear panel data model is that seasonality can be added. Day of the week effects are considered in the first model for both countries. This addition goes in line with (Gugler et al., 2021).

Ordinary Least Squares for Panel Data

To test the first hypothesis, the panel data linear model is used. It is based in the models presented by Drukker (2003), Metcalf and Stock (2020), and Gugler et al. (2021). It is as follows:

(ii) $y_{it} = \alpha + X_{it}\beta_1 + y_{t-1}\beta_2 + W_j\delta_1 + Z_i\delta_2 + \varepsilon_i t$
where $i \in \{1, 2, ..., N\}, t \in \{1, 2, ..., T_i\}, j = 7$

In the equation (ii), y_{it} represents the dependent variable. In the analysis of this paper, that represents the CO_2 emissions from Coal and Gas of the UK and Coal, Gas and Lignite of Germany, each in one independent equation. X_{it} represents a matrix of independent variables, which are time-varying. The size of the matrix is $(8xK_1)$, because eight exogenous variables are used in the analysis. y_{t-1} represents the past returns of the dependent variable. This addition was based in the paper presented by (Metcalf & Stock, 2020). W_t represent a matrix of time-invariant covariates. The size of the matrix is $(1x7)$. It represents the day of the week effect, which goes from 1 to 7, where 1 represents Sunday and 7 Saturday. (Gugler et al., 2021) considered daily and monthly effects in their analysis. The parameters α, β_1, and δ_1 represent the relationship between the dependent and each of the independent variables. $\varepsilon_i t$ is the idiosyncratic error. All variables used in the equation (ii) are logarithmic returns calculated according to the equation (i).

Differences in Differences (DD) for Panel Data

Table 3: Summary of Statistics of daily data

Germany

	Mean	Std. Dev.	Min	p25	p50	p75	Max
CO_2 emissions of Gas (thousands of tons)	51	26	9	29	47	70	134
CO_2 emissions of Coal (thousands of tons)	126	80	20	56	108	183	343
CO_2 emissions of Lignite (thousands of tons)	345	98	90	281	377	420	496
Electricity generation by Nuclear (GWh)	189	31	99	160	188	216	247
Electricity generation by Solar (GWh)	113	73	5	44	108	173	290
Electricity generation by Wind Offshore (GWh)	60	36	1	29	58	88	145
Electricity generation by Wind Onshore (GWh)	258	188	13	114	207	351	914
EUA (€ /tCO_2)	17.9	8.4	4.3	7.9	20.6	25.1	33.3
Electricity demand (GWh)	1409	193	885	1265	1420	1551	1890
DAX (€)	12285	833	8442	11890	12382	12902	13790
Coal-to-Gas Ratio (1, used for both countries)	0.67	0.2	0.44	0.55	0.64	0.71	1.82

United Kingdom

	Mean	Std. Dev.	Min	p25	p50	p75	Max
CO_2 emissions of Gas (thousands of tons)	122	38	3	93	122	149	242
CO_2 emissions of Coal (thousands of tons)	27	37	0	3	13	35	219
Electricity generation by Nuclear (GWh)	154	26	64	139	155	174	206
Electricity generation by Solar (GWh)	30	20	1	13	27	45	81
Electricity generation by Wind Offshore (GWh)	53	34	2	25	46	75	151
Electricity generation by Wind Onshore (GWh)	76	39	1	44	72	105	194
UK CPS (€ /tCO_2)	38.3	8.4	25.1	28.3	40.9	45.6	53.1
Electricity demand (GWh)	714	115	243	638	703	784	1040
FTSE 100 (€)	7071	588	4994	6940	7291	7454	7877

The DD model is useful to see the effect that a treatment (i.e. a government policy) had in a group versus another (Angrist & Krueger, 1999). It has been applied to test policies in economics (Angrist & Krueger, 1999), education (Schwerdt & Woessmann, 2020), and carbon price (Abrell et al., 2021) and (Marion, 2019). The method is called differences in differences because it takes a double difference. It is the difference of the outcome's estimation without the government policy versus the outcome after the policy implementation (Angrist & Krueger, 1999). For this study, two periods and two groups are required (Schwerdt & Woessmann, 2020). In the first period, none of the groups is affected by the treatment. In the second period, only one of them is. The group that is not exposed, is called the control group. In this paper, that is the United Kingdom. Even though the UK was also affected by the MSR, its trend of CO_2 emissions did not change after the policy implementation. On the other hand, the trend of CO_2 emissions in Germany changed after the MSR. That condition allows the use of the DD model in this comparison (Schwerdt & Woessmann, 2020). The fact that the UK had a high enough carbon price before the policy explains this difference. Gugler et al. (2021) found that in the UK, a carbon price above 38€ /ton was less effective because it affected fewer coal-fired power plants. Thus, in the UK, the MSR had almost no impact, contrary to its effect in Germany.

As mentioned before, the DD method is a variation of the OLS for panel data. It is specified as follows:

(iii) $y_t = \alpha + T_t\beta_1 + S_t\beta_2 + (T_t * S_t)\beta_3 + Z\delta_1 + \varepsilon_t$
where $t \in \{1, 2, ..., T\}$

In the equation (iii), y_t represents the total CO_2 emissions of each country. T_t is a dummy variable that represents the treatment. It takes the value of 1 during the years 2019 and 2020 when the MSR was active. T_t is a dummy variable that represents the country affected by the policy. It takes the value of 1 for Germany and 0 for the UK. Therefore the $T_t * S_t$ represents Germany when the MSR was active. As in the OLS panel data, the parameters $\alpha, \beta_1, \beta_2, \beta_3$ and δ_1 represent the relationship between the dependent and each of the independent variables. Finally, $\varepsilon_i t$ is the idiosyncratic error. Figure 7 shows the intuition behind the model. In the x-axis is represented by T_t, where the years 2017-2018 take the value of 0 and 2019-2020 the value of 1. The y-axis represents the daily average of CO_2 emissions for each country. Finally, the bold grey line represents the differences in differences. The model optimization allows to determine 1) the size of the differences in differences and 2) its significance.

In this paper, specification tests will be carried out. In the incoming paragraphs, an explanation of each of them will be presented.

Analysis of Variance (ANOVA)

To get the right model, each exogenous variable must have a self-explanatory power. That means that each of them must be independent of the other. Otherwise, it should be eliminated from the model. The ANOVA test allows comparing

Table 4: Summary of Statistics of daily returns

Germany

	Mean	Std. Dev.	Min	p25	p50	p75	Max
CO_2 emissions of Gas (thousands of tons)	0.00	0.34	-1.26	-0.19	-0.01	0.15	1.16
CO_2 emissions of Coal (thousands of tons)	-0.00	0.49	-1.62	-0.26	-0.02	0.20	1.75
CO_2 emissions of Lignite (thousands of tons)	-0.00	0.22	-1.10	-0.06	-0.00	0.07	0.96
Electricity generation by Nuclear (GWh)	0.00	0.06	-0.40	-0.01	-0.00	0.01	0.34
Electricity generation by Solar (GWh)	0.00	0.39	-1.66	-0.21	0.00	0.20	1.68
Electricity generation by Wind Offshore (GWh)	-0.00	0.81	-4.00	-0.42	-0.01	0.42	3.29
Electricity generation by Wind Onshore (GWh)	-0.00	0.66	-2.42	-0.43	-0.01	0.42	2.36
EUA (€ /tCO_2)	0.00	0.02	-0.19	-0.01	0.00	0.01	0.13
Electricity demand (GWh)	-0.00	0.10	-0.35	-0.06	-0.01	0.05	0.33
DAX (€)	0.00	0.01	-0.13	-0.00	0.00	0.01	0.24
Coal-to-Gas Ratio (1, used for both countries)	-0.00	0.03	-0.34	-0.01	0.00	0.01	0.24

United Kingdom

	Mean	Std. Dev.	Min	p25	p50	p75	Max
CO_2 emissions of Gas (thousands of tons)	0.00	0.30	-2.28	-0.16	-0.01	0.18	3.21
CO_2 emissions of Coal (thousands of tons)	0.07	1.04	-9.70	-0.28	0.03	0.45	7.24
Electricity generation by Nuclear (GWh)	-0.00	0.05	-0.35	-0.01	0.00	0.02	0.31
Electricity generation by Solar (GWh)	0.00	0.60	-2.55	-0.30	0.01	0.30	2.44
Electricity generation by Wind Offshore (GWh)	-0.00	0.64	-2.38	-0.40	-0.01	0.38	2.65
Electricity generation by Wind Onshore (GWh)	-0.00	0.55	-1.97	-0.33	-0.00	0.33	4.01
UK CPS (€ /tCO_2)	0.00	0.01	-0.10	-0.00	0.00	0.00	0.06
Electricity demand (GWh)	-0.00	0.10	-0.71	-0.05	-0.01	0.04	0.78
FTSE 100 (€)	-0.00	0.01	-0.12	-0.00	0.00	0.00	0.09

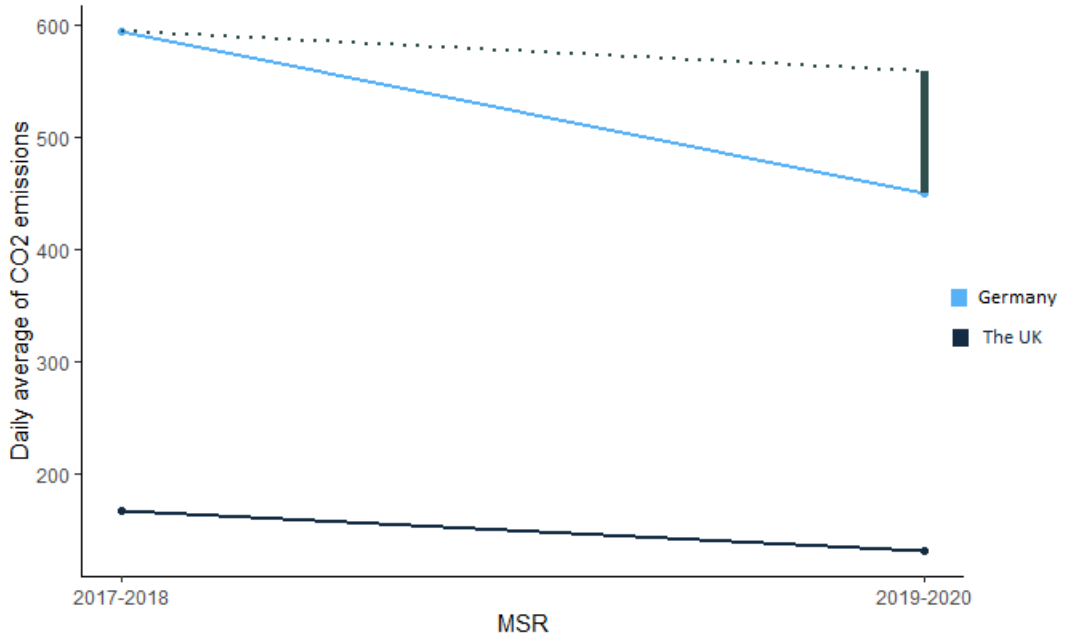


Figure 7: DD between Germany and the UK after the MSR implementation

Figure created by the author

models, with the objective to get the best fit (Faraway, 2002). All the models presented in this paper have approved the

ANOVA test. In the case of the UK, the test eliminated the Coal-to-Gas price ratio and the FTSE 100, because both were considered not relevant to estimate the CO_2 emissions of coal nor gas. In the case of Germany, the returns of the DAX were eliminated from the three models. However, the Coal-to-Gas ratio was considered significant to explained the CO_2 emissions of Lignite.

Quantile-Quantile Plot for Residuals

The quantile-quantile (Q-Q) plot is a graphical tool. It shows whether the results of the model are normally distributed or not (University of Virginia, 2015). It plots two sets of quantiles together. In the case that both are normally distributed, the points will be located near the line (University of Virginia, 2015). For the models used in this paper, all the residuals showed a normal distribution. The results of the Q-Q plots can be found in the Appendix.

Durbin-Watson Test for the Residuals

The residuals of the model employed in this paper must be independent. This assumption allows a valid interpretation of the model, even if the observations used are serially correlated (Durbin & Watson, 1950). The Durbin-Watson test checks two assumptions: 1) if the error is independently distributed of the exogenous variables with a mean of zero and a constant variance, and 2) if successive errors are independently distributed of one another (Durbin & Watson, 1950). Table 5 shows the Durbin-Watson results for the models computed in this paper. All the models accept the null hypothesis that the residuals are not autocorrelated at a significance level of 99%.

Table 5: Results of Durbin-Watson Test

Germany	
	p-value
CO_2 emissions of Gas	1
CO_2 emissions of Coal	0.9998
CO_2 emissions of Lignite	0.9909
The United Kingdom	
	p-value
CO_2 emissions of Gas	1
CO_2 emissions of Coal	0.9884
Differences in Differences	
	p-value
Panel model	1

The paper will continue as follows: In the next section, the empirical findings are discussed. Then, the last section of the document concludes the study with the main findings and future research directions.

7. Empirical results

The interpretation of the results in the tables (6-8) is as follows. First, the adjusted R-squared is shown. That number reveals how much of the mean return of CO_2 emissions per fossil fuel per country is explained by the independent variables. Second, the standard error shows the average distance between the observed values and the ones estimated by the model or coefficient. A smaller value is always preferable, but it is expressed in relative terms. This means that a larger coefficient will tend to have a larger standard error. The degrees of freedom are the difference between the number of observations and the independent variables. Also, the coefficient of each variable is shown. Since it is expressed in returns, the coefficient indicates how the variation of each variable affects the CO_2 emissions. The t-value shows the significance level of each estimator. The asterisks are based on the confidence levels that are shown at the end of the tables. Finally, it is worth mentioning that only significant variables are shown in the models since we follow the ANOVA test.

After the estimation of the panel OLS and DD models with the software R, the results obtained are shown in tables 6 to 8. For the UK, the results are shown in Table 6. Following the ANOVA test, the Coal-to-Gas price ratio, the FTSE 100, the day of the week effect of Sunday, Wednesday, and Thursday were eliminated, because they were not significant. For the UK, the positive coefficient of days Monday and Tuesday shows that electricity generated by coal increased in those days. While, on Friday and Saturday, it decreased. Both are explained by the interaction of cost and demand. Since coal is the most expensive fossil fuel in the merit-order curve for the UK, its production increases only when there is a jump in demand (i.e. on Mondays and Tuesdays). On the side of gas, its generation decreases on Mondays but increases on Saturdays. Nevertheless, both effects in natural gas are weak. Finally, the CO_2 emissions from coal and gas are explained by 24.9% and 78.8%, respectively.

Table 7 shows the results obtained for Germany. The CO_2 emissions from coal, gas, and lignite are explained by 66.2%, 66.2%, and 52.4%, respectively. Following the ANOVA test, the DAX, and the day of the week effect for Wednesday and Thursday were eliminated. For all fossil fuels, there is a negative coefficient on Sundays, Fridays, and Saturdays. As in the UK, the position of these fossil fuels on the merit order explains this. Since they are more expensive, when the demand is lower, renewable energies cover the demand. On Mondays and Tuesdays, these fossil fuels are used since the demand is higher. The model that estimates the CO_2 emissions from lignite has a different output. The lagged returns, the electricity produced by solar and wind offshore, and the electricity load were removed for the estimation. However, the Coal-to-Gas price ratio was considered significant and reveals that the average of CO_2 emissions from lignite decreases 0.24% when coal gets more expensive than gas by 1%. Finally, the results of the DD model are presented in table 8. For this estimation, the ANOVA method eliminated the market indices. The DD model explains 97.9% of the difference in CO_2 emissions

between Germany and the UK. Despite the difference in the R-squared, the tests presented in the methodology validate all models.

The rest of the results will be explained in detail in the next paragraphs. The analysis will comprise the hypothesis 1 and 2, nuclear, solar, and wind energy.

7.1. Hypothesis 1

The findings provide empirical evidence that there is a link between the carbon price in the UK and the CO_2 emissions. This link is negative for coal and positive for gas, following the results found by Gugler et al. (2021), Abrell et al. (2021), Marion (2019), and Wilson and Staffell (2018). These authors concluded that the UK CPF has been effective in reducing the CO_2 emissions in the UK. Moreover, that the UK CPF promoted the switch from coal to gas, because its coefficient (impact) is negative on coal, but positive on gas. Table 6 shows the coefficients and its significance. For coal, the magnitude of its coefficient is large, as well. It means that an increase of 1% on the carbon price will decrease the CO_2 emissions from coal by 4%. On the side of gas, an increase of 1% on the carbon price will increase its CO_2 emissions by 0.5%. This difference in magnitude makes sense when the increase in renewable and nuclear energy is included. Because both increased. The differences between this paper and those mentioned above are 1) the period analyzed was from 2017 to 2020 2) this paper took national electricity generation variables (not production per power plants), and 3) the exclusion of electricity produced by nuclear energy. The latter proved to be a significant variable across both countries and fossil fuels.

Conversely, the results of Germany show that the EU ETS does not explain its CO_2 emissions during the period analyzed. As observed in table 7, the carbon price was found not significant across the three fossil fuels analyzed. This result is not directly comparable with the literature described in this paper, because no studies were analyzing the impact of the EU ETS in Germany. However, the results agree partially with Gugler et al. (2021). The authors compared the effectiveness of carbon pricing and renewables subsidies in the UK and Germany, respectively. They concluded that carbon pricing has been more effective because the British price was high enough. The latter did not happen in Germany during all the period analyzed. Figure 5 shows that the average price of carbon in Germany was 5.8€ /tCO_2 in 2017. According to Gugler et al. (2021), 8€ /tCO_2 is the minimum effective carbon price for Germany. However, the carbon price oscillated between 16 € /tCO_2 and 24 € /tCO_2 during three-quarters of the data analyzed. Three factors can explain this difference. First, the period analyzed. This paper analyzed data from January 2017 to December 2020, while Gugler et al. (2021) took data from January 2017 to June 2018. Second, they took electricity generation per power plant, while in this paper was at national levels. Third, the inclusion of nuclear energy.

7.2. Hypothesis 2

Since the DD model is expressed in differences, some clarifications have to be mentioned to interpret the results correctly. First, a negative coefficient means that the reduction of CO_2 emissions was larger in Germany than in the UK. This is validated by Figures 6 and 7. Both figures exclude the alternative interpretation that the UK had reduced its emissions more. Second, three conditional variables are considered. First, the variable *Germany (2017-2020)* is not a difference. It only considers the CO_2 emissions of Germany during the whole sample. Second, *MSR (2019-2020)* shows the joint average of CO_2 emissions of both Germany and the UK. Third, *Germany (2019-2020)* shows the difference in CO_2 emissions of Germany and the UK, when the MSR was active. The latter is the one of interest for this paper. Nevertheless, it is complemented with both variables: *Germany (2017-2020)* and *MSR (2019-2020)*.

The findings provide empirical evidence that there is a link between the introduction of the MSR and the CO_2 emissions in Germany. This link is negative and statistically significant. As shown in Table 8, only because of the MSR implementation, Germany decreased its daily CO_2 emissions by 39.5 tonnes in comparison with the UK. However, two additional effects resulted. First, the MSR also impacted the CO_2 emissions of the UK. On average, each country reduced its emissions by 17.7 during this time. Second, Germany alone reduced its average daily CO_2 emissions by 178.6 during the years from 2017 to 2020. In summary, the MSR helped the CO_2 emissions of both Germany and the UK, but its impact on Germany was larger. This happened during years where the average carbon price of the UK and Germany, was 50 € /tCO_2, and 25 € /tCO_2, respectively. The latter goes in line with Gugler et al. (2021). The authors resolved that in the UK when the carbon price was above 38 € /tCO_2, its marginal benefit started to decline. The latter happened during the years 2019 and 2020. As shown in Figure 5, the average carbon price in the UK changed from 36 € /tCO_2 in 2018 to 45 € /tCO_2 in both 2019 and 2020. Finally, the coefficient *Germany (2019-2020)* is significant and negative despite the development of the Covid-19 crisis that starts in February 2020. This result agrees with the conclusions of Gerlagh et al. (2020), who found that the MSR was a good stabilizer during the Covid-19 crisis.

The control variables use in the DD model increase its statistical power and allow to see how other factors affect the difference in CO_2 emissions. The carbon price shows that Germany emitted on average an additional of 1.9 CO_2 tonnes in comparison with the UK. This makes sense because the carbon price of the UK was higher than the one in Germany during the years analyzed. The generation of electricity by renewables and nuclear sources has a negative coefficient. It means that Germany reduced an additional of -0.7 tonnes of CO_2 emissions because of its generation of clean electricity. This is supported by the fact that during the years analyzed, Germany increased its share of clean electricity by 10.86% compared with 8.64% of the UK (Our World in Data, 2021). Finally the positive coefficient of electricity generation shows

Table 6: United Kingdom - Results of the panel OLS on CO_2 emissions

	Coal			Gas		
Adj. R-squared	24.9%			78.8%		
Residual st. error	0.89			0.14		
Degrees of freedom	1443			1445		
	Estim.	Std. Err.	t-val	Estim.	Std. Err.	t-val
Lagged (t-1)	-0.23	0.02	-9.33***	NA	NA	NA
Nuclear	-1.09	0.55	-1.97**	-0.70	0.09	-8.02***
Solar	-0.07	0.04	-1.94*	-0.05	0.01	-7.50***
Wind Onsh.	-0.24	0.06	-4.17***	-0.15	0.01	-16.49***
Wind Offsh.	-0.17	0.05	-3.51***	-0.11	0.01	-14.62***
Carbon Price	-3.93	1.97	-2.0**	0.53	0.3	0.09*
Load	2.57	0.38	6.73***	2.55	0.06	42.63***
Day Week: Monday	0.38	0.08	4.68***	-0.06	0.01	-4.71***
Day Week: Tuesday	0.15	0.06	2.32**	0.01	0.01	0.55
Day Week: Friday	-0.11	0.06	-1.82*	0.01	0.01	1.11
Day Week: Saturday	-0.29	0.07	-3.83***	0.05	0.01	4.27***
Confidence Levels:						
*: 90%						
**: 95%						
***: 99%						

Table 7: Germany - Results of the panel OLS on CO_2 emissions

	Coal			Gas			Lignite		
Adj. R-squared	66.2%			66.2%			52.4%		
Residual st. error	0.28			0.20			0.15		
Degrees of freedom	1443			1443			1447		
	Estim.	Std. Err.	t-val	Estim.	Std. Err.	t-val	Estim.	Std. Err.	t-val
Lagged (t-1)	-0.05	0.02	-2.62***	-0.08	0.02	-3.94***	NA	NA	NA
Nuclear	0.71	0.14	5.06***	0.68	0.10	7.04***	1.30	0.07	17.73***
Solar	-0.11	0.02	-5.77***	-0.08	0.01	-6.00***	NA	NA	NA
Wind Onsh.	-0.38	0.02	-21.5***	-0.23	0.01	-18.88***	-0.11	0.01	-18.10***
Wind Offsh.	-0.06	0.01	-5.23***	-0.01	0.01	-1.74*	NA	NA	NA
Carbon Price	-0.17	0.3	-0.57	0.09	0.21	0.43	-0.05	0.16	-0.31
Load	2.04	0.15	13.53***	0.68	0.10	6.53***	NA	NA	NA
Coal-to-Gas	NA	NA	NA	NA	NA	NA	-0.24	0.13	-1.90*
Day Week: Sunday	-0.09	0.02	-3.69***	-0.12	0.02	-7.54***	-0.05	0.01	-5.18***
Day Week: Monday	0.32	0.03	9.93***	0.32	0.02	14.62***	0.17	0.01	15.81***
Day Week: Tuesday	0.10	0.02	4.10***	0.10	0.02	6.13***	0.03	0.01	3.46***
Day Week: Friday	-0.07	0.02	-3.48***	-0.04	0.01	-3.18***	-0.02	0.01	-2.07**
Day Week: Saturday	-0.24	0.03	-9.42***	-0.27	0.02	-15.05***	-0.13	0.01	-12.51***
Confidence Levels:									
*: 90%									
**: 95%									
***: 99%									

that Germany produced more CO_2 emissions in comparison with the UK. As explained in section 3.3.2, the UK is one one of the IEA's leading countries in energy efficiency per GDP (IEA, 2019).

In summary, the DD model validates the second hypothesis. It allows us to conclude that the MSR acted as a carbon price floor for Germany. The model shows that the MSR impacted the CO_2 emissions of Germany in a larger and significant way when compared with the UK. Moreover, this happened in times where the economic crisis Covid-19 took place.

7.3. Individual analysis of relevant variables

The goal of this paper is to determine whether the CPF in the UK has been more effective in tackling the CO_2 emissions. However, both policies, the EU ETS and the UK CPF aim to

Table 8: Germany and the UK - Results of DD model

Adj. R-squared	97.9%		
Residual st. error	33.08		
Degrees of freedom	2911		
	Estim.	Std. Err.	t-val
Intercept	-179.4	5.4	-33.2***
Germany (2017-2020)	-178.6	4.2	42.7***
MSR (2019-2020)	-17.7	2.5	-7.1***
Carbon Price	1.9	0.1	14.9***
Renewables + Nuclear	-0.7	0.0	-151***
Electricity Generation	0.7	0.0	153.9***
Germany (2019-2020)	-39.5	2.5	-15.9***
Confidence Levels:			
*: 90%			
**: 95%			
***: 99%			

increase investments in renewable energies, too. Therefore, an analysis of the development of nuclear, wind, and solar electricity will be carried out. The coefficients from Tables 6-8 will be used in this analysis.

7.3.1. Nuclear energy

As explained in sections 3.3 and 4.1, the inclusion of nuclear energy is meaningful because 1) it represents more than 10% of the electricity produced in both Germany and the UK, and 2) Germany will phase out nuclear energy in 2022, while the UK considers it as fundamental to meet its environmental goals. Accordingly, the relationships between nuclear electricity and CO_2 emissions differ per country.

In the UK, there is a negative and significant link between nuclear electricity and CO_2 emissions from both coal and gas. Thus, the inclusion of nuclear electricity as a relevant variable is validated. For coal, it means that an increase of 1% of electricity from nuclear decreases CO_2 emissions from coal by 1.09%. For gas, an increase of 1% of electricity from nuclear decreases its CO_2 emissions by 0.7%. However, during the years analyzed (2017-2020), electricity generation from coal, gas, and nuclear decreased by 5%, 4.5%, and 4.1%, respectively. How can this relationship be explained? According to IEA (2019), despite the UK Government supports nuclear, the technology faces challenges. In June 2018, the UK Government released the Nuclear Sector Deal. The objective of the policy was to deliver affordable and reliable nuclear power, by increasing investments in innovation, supply-chain, and construction of new Nuclear Power Plants (NPPs) (IEA, 2019). However, eight nuclear reactors will be shut down in 2023, and three NPPs have not started to be constructed (IEA, 2019). In summary, the negative coefficients shown in Table 6 indicate that nuclear energy has replaced the electricity generated by coal and gas in the UK.

Conversely, there is a positive and significant relationship between nuclear electricity and CO_2 emissions from coal, gas, and lignite. As in the UK, nuclear electricity is a variable that effectively explains CO_2 emissions. However, in Germany, additional nuclear electricity increases them. Table 7 shows these coefficients. For lignite, the relationship is the strongest one. It shows that an increase of 1% of electricity from nuclear increases CO_2 emissions from lignite by 1.3%. For both coal and gas, the increment is 0.7%. The coefficients indicate that nuclear energy generates electricity in parallel with coal, gas, and lignite. Demand peaks could be an explanation, since they are covered by the most expensive sources, which are fossil fuels. Moreover, the positive coefficients in the model concur with the positive correlations shown in the Table 2. As mentioned in section 3.3.1, Germany will phase out nuclear energy in 2022. Also, it plans to cover the gap left by nuclear with natural gas and imports of electricity (IEA, 2021).

7.3.2. Solar and wind energy

The relationship of electricity generated by solar and wind (onshore+offshore) and CO_2 emissions is negative and significant across both countries and fossil fuels, excluding lignite. Both findings go in line with the development that Germany and the UK have achieved in solar and wind energy. Currently, solar generation in Germany is one of the highest in the world (Clean Energy Wire, 2020). Also, it counts with the support of the Government as part of the *Energiewende* policy, and it is highly approved by citizens (Clean Energy Wire, 2020). On the other hand, the UK is a leader in wind energy. The country is the world leader in wind offshore by installed capacity (Renewable UK, 2021). The latter is confirmed by its coefficient shown in Table 6, which is the highest across countries.

Wirth (2021) agree that despite PV and wind energy have increased their efficiency, are not capable of replacing fossil fuels in the near future. The lack of electricity storage and weather conditions are among the principal reasons. Gates (2021) agrees with this point and adds another point. Gates (2021) states that the power generated by PV and wind per square meter is limited. On average, fossil fuels generated 500-10,000 watts per square meter. Nuclear between 500-

1,000 (Gates, 2021). However, solar energy generates only between 5-20 watts per square meter, and wind 1-2 (Gates, 2021). About this, Wirth (2021) states that, in Germany, it is possible to have enough space for PV if they are integrated. Integrated PV are part of buildings, automobiles, in parks, roads, and on top of agriculture plantations (Wirth, 2021).

The paper will continue as follows: In the next section, the conclusions of this study and future research directions are presented.

8. Summary and Conclusions

This paper analyzed two research questions. First, the impact of carbon pricing on tackling CO_2 emissions in Germany and the United Kingdom. Second, whether the Market Stability Reserve introduced in 2019 acted as a Carbon Price Floor (CPF) for Germany.

By using an Ordinary Least Squares (OLS) model for panel data, it was determined that the United Kingdom was more effective in combating CO_2 emissions due to its CPF policy. This result supports existing evidence of Gugler et al. (2021), who established that the carbon policy of the UK was more effective in reducing CO_2 emissions than the subsidies that Germany gave to boost renewable technologies. Moreover, authors (see (Abrell et al., 2021; Marion, 2019)), who studied the impact of the CPF in the UK, determined that the policy was effective in reducing CO_2 emissions. Other authors (see (Edenhofer et al., 2017; Gerlagh et al., 2020)), who studied the EU ETS in Europe, resolved that a CPF would improve the effectiveness of the EU ETS.

To test the second hypothesis, a model of Differences in Differences (DD) for panel data was employed. The model determined that the MSR reduced the CO_2 emissions of both countries. However, its impact increased significantly in Germany, enabling its comparison with the CPF of the UK. This result endorsed the findings of Gugler et al. (2021), who determined that when the British carbon price was above 38 € /tCO_2, its marginal benefit started to decline. The British carbon price exceeded these levels during 2019 and 2020 when the MSR was in operation. Thus, the MSR was significantly more effective in tackling the CO_2 emissions of Germany.

Previous research has mainly focused on evaluating the impact of carbon pricing in CO_2 emissions (see (Abrell et al., 2021; Gerlagh et al., 2020; Gugler et al., 2021)) but has excluded one of these two factors: the influence of nuclear energy, and the impact of the MSR in the EU Emissions Trading Scheme (EU ETS). Both factors are considered in this research because nuclear energy is a carbon-free source and the MSR acted as a market stabilizer. Accordingly, the results of this study demonstrate that both factors are relevant for the analysis.

Nuclear energy proved to be a relevant factor of the CO_2 emissions across countries and fossil fuels. In the UK, nuclear has a negative relationship with CO_2 emissions of both coal and gas. It means that nuclear acted as a substitute for fossil fuels. Since the British Government supports nuclear energy, the finding makes sense. Inversely, in Germany, nuclear energy has a positive link with CO_2 emissions of all fossil fuels. It denotes that when Germany produced nuclear energy, it also increased its production from coal and gas by a factor of 0.7, approximately. The link with lignite was stronger, of 1.3x. In summary, nuclear acted as a complementary for fossil fuels. Since Germany has reduced its nuclear generation due to the phase-out in 2022, both fossil fuels and nuclear may have been necessary to fulfill its electricity demand at the same time. Consequently, the link that each country has between nuclear energy and CO_2 emissions corresponds to their opposing views on nuclear.

Since the introduction of the EU ETS in 2005, the EU Commission has taken feedback and has actively improved its policy. The MSR established in 2019 was one of the improvements. The MSR controls the supply of European Allowances (EUAs) and avoided a price crash during the Covid-19 crisis. Despite the MSR was active for both countries, this paper has proved that it affected mostly German's CO_2 emissions. The DD model resolved that the MSR accounted for a daily reduction of Germany's CO_2 emissions by 39.5 tonnes in comparison with the UK. Currently, the discussion of whether the creation of a CPF for the EU ETS is open (see (Gerlagh et al., 2020)). Annalena Baerbock, the leader of the German Green Party, announced that, if elected, her party will raise the carbon price to 60 € /ton by 2023 (NTV, 2021). The finding that the MSR may have acted as a CPF for Germany, adds value to the existing literature and debate.

However, the promotion of renewable energies is the final goal of carbon pricing. Therefore, this paper has included the electricity generation of solar and wind energy in the analysis. The finding is that electricity generation from renewables has reduced CO_2 emissions across countries and fossil fuels. For Germany, solar energy and wind onshore have displaced the highest share of coal and gas. Lignite has been significantly affected only by wind onshore. For the UK, both wind onshore and offshore were the most significant sources. The results go in line with each country's development. Germany ranked 4th in the world for its photovoltaic installed capacity (Clean Energy Wire, 2020). While the UK has the largest capacity of offshore wind in the world (Renewable UK, 2021).

Given the results of this paper and prior findings, it is clear that carbon pricing has reduced CO_2 emissions and has promoted the development of renewable energies. Also, these findings demonstrate that the reforms taken by the EU Commission to consolidate the EU ETS are effective, too. Still, it is important to mention that in February 2018, the EU Commission approved a reform of the EU ETS for phase 4 (2021 - 2030). The period analyzed in this paper excludes the latest reform. Beck and Kruse-Andersen (2020) run simulations until 2125 to estimate the effects of this reform. They concluded that the new MSR was more effective in the short and long run. The authors stated that the new MSR was being affected by the market demand on EUAs, and thus it was not being set by the EU Commission. The new MSR gives space for further research.

There are some limitations to this research. First, the focus of the research is on analyzing the impact of carbon pricing on CO_2 emissions. Investments in renewables and energy efficiency improvements are excluded. The inclusion of electricity load relates partially to energy efficiency, but it is not the same. Second, electricity imports are excluded from the analysis. Marion (2019) points out that it is unlikely that the UK has increased its electricity imports because the UK is not well interconnected. The same cannot be said about Germany.

This study opens two discussion points in carbon pricing. First, the effectiveness of the MSR as a market stabilizer, and therefore as a CPF for Germany. Even though the global economy has not recovered from the Covid-19 crisis, the EU ETS price has kept its uptrend in 2021. From 4th January to 11th August, the EU ETS price has increased from 33.7 € /ton to 57.8 € /ton (Ember Climate, 2021). The latest reform mentioned in the previous paragraph may have improved the effectiveness of the EU ETS. Second, the consideration of national energy policies when two countries are compared. The inclusion of nuclear energy in the model allows us to see that the effectiveness of the UK CPF, when compared to Germany, is positively biased by its nuclear policy.

Despite 191 parties (including the EU) signed the 2015 Paris Agreement, the EU and the UK are pioneers in implementing a carbon market (The United Nations, 2021). These complementary schemes have been analyzed in this study and the results lead to two recommendations for countries that are adopting an ETS (i.e. South Korea and China). First, that the MSR could be as effective as a CPF. Second, to assess their individual country policies (i.e. nuclear policy) when designing an ETS.

References

Abrell, J., Faye, A. N., & Zachmann, G. (2011). Assessing the impact of the EU ETS using firm level data [Working Papers of BETA]. (2011-15). Retrieved from https://ideas.repec.org/p/ulp/sbbeta/2011-15.html

Abrell, J., Kosch, M., & Rausch, S. (2021). How effective is carbon pricing? emissions and cost impacts of the uk carbon tax.

Anderson, B., & Di Maria, C. (2011). Abatement and allocation in the pilot phase of the eu ets. *Environmental and Resource Economics*, *48*, 83-103. doi: https://doi.org/10.1007/s10640-010-9399-9

Angrist, J. D., & Krueger, A. B. (1999). Empirical strategies in labor economics. In *Handbook of labor economics* (Vol. 3, pp. 1277–1366). Elsevier.

BBC. (2012). *Gas fracking: Ministers approve shale gas extraction.* Retrieved 2021-06-28, from https://www.bbc.com/news/uk-20707574

Beck, U., & Kruse-Andersen, P. K. (2020). Endogenizing the cap in a cap-and-trade system: Assessing the agreement on eu ets phase 4. *Environmental and Resource Economics*, *77*, 781-811. Retrieved from https://doi.org/10.1007/s10640-020-00518-w doi: 10.1007/s10640-020-00518-w

Bel, G., & Joseph, S. (2015). Emission abatement: Untangling the impacts of the eu ets and the economic crisis. *Energy Economics*, *49*, 531-539. Retrieved from https://www.sciencedirect.com/science/article/pii/S0140988315001061 doi: https://doi.org/10.1016/j.eneco.2015.03.014

Bourn, J. (2004). *The uk emissions trading scheme: A new way to combat climate change.*

Bruggeman, J., & Gonenc, H. (2013). Emissions trading and stock returns: Evidence from the european steel and combustion industries. *Energy Economics and Financial Markets*. doi: https://doi.org/10.1007/978-3-642-30601-3_7

Carbon Brief. (2019). *Guest post: Why german coal power is falling fast in 2019.* Retrieved 2021-05-24, from https://www.carbonbrief.org/guest-post-why-german-coal-power-is-falling-fast-in-2019

Clean Energy Wire. (2020). *Solar power in germany – output, business & perspectives.* Retrieved 2021-08-11, from https://www.cleanenergywire.org/factsheets/solar-power-germany-output-business-perspectives

Clean Energy Wire. (2021). *The history behind germany's nuclear phase-out.* Retrieved 2021-08-03, from https://www.cleanenergywire.org/factsheets/history-behind-germanys-nuclear-phase-out

Declercq, B., Delarue, E., & D'haeseleer, W. (2011). Impact of the economic recession on the european power sector's co2 emissions. *Energy Policy*, *39*(3), 1677-1686. Retrieved from https://www.sciencedirect.com/science/article/pii/S0301421510009432 doi: https://doi.org/10.1016/j.enpol.2010.12.043

Deutsche Welle. (2019). *German wind energy stalls amid public resistance and regulatory hurdles.* Retrieved 2021-06-28, from https://www.dw.com/en/german-wind-energy-stalls-amid-public-resistance-and-regulatory-hurdles/a-50280676

Dhamija, A. K., Yadav, S. S., & Jain, P. (2018). Volatility spillover of energy markets into EUA markets under EU ETS: a multi-phase study. *Environmental Economics and Policy Studies*, *20*(3), 561–591. Retrieved from https://doi.org/10.1007/s10018-017-0206-5 doi: 10.1007/s10018-017-0206-5

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *The stata journal*, *3*(2), 168–177.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika*, *37*(3/4), 409–428.

Edenhofer, O., Flachsland, C., Wolff, C., Schmid, L. K., Leipprand, A., Koch, N., … Pahle, M. (2017). *Decarbonization and EU ETS Reform: Introducing a price floor to drive low-carbon investments* (1st ed.).

Eichhammer, W., Friedrichsen, N., Healy, S., & Schumacher, K. (2018). Impacts of the allocation mechanism under the third phase of the european emission trading scheme. *Energies*, *11*. doi: 10.3390/en11061443

Ellerman, A., Convery, F., & de Perthuis, C. (2010). *Pricing carbon - the european union emissions trading scheme*. New York.

Ember Climate. (2021). *Eua (eu ets) futures prices.* Retrieved 2021-08-12, from https://ember-climate.org/data/carbon-price-viewer/

European Commission. (2007). *Emissions trading: strong compliance in 2006, emissions decoupled from economic growth.* Retrieved 2021-07-25, from https://ec.europa.eu/commission/presscorner/detail/en/ip_07_776

European Commission. (2012). *Emissions trading: Commission presents options to reform the european carbon market.* Retrieved 2021-07-25, from https://ec.europa.eu/commission/presscorner/detail/en/ip_12_1208

European Commission. (2013). *Climate action: Commission proposes ratification of second phase of kyoto protocol.* Retrieved 2021-05-24, from https://ec.europa.eu/commission/presscorner/detail/en/ip_13_1035

European Commission. (2014). *State aid: Commission approves uk compensation for indirect costs of carbon price floor.* Retrieved 2021-05-24, from https://ec.europa.eu/commission/presscorner/detail/en/ip_14_577

European Commission. (2015a). *Eu ets handbook.*

European Commission. (2015b). *Speech by commissioner arias cañete at the lisbon council "towards an effective energy union".* Retrieved 2021-05-24, from https://ec.europa.eu/commission/presscorner/detail/en/speech_15_4439

European Commission. (2017). *Eu emissions trading system: landmark agreement between parliament and council delivers on eu's commitment to turn paris agreement into reality.* Retrieved 2021-05-24, from https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_17_4501

European Commission. (2020). *Opening remarks by evp timmermans and commissioner simson at the press conference on the 2030 climate target plan.* Retrieved 2021-05-24, from https://ec.europa.eu/commission/presscorner/detail/en/speech_20_1692

European Commission. (2021a). *Eu emissions trading system (eu ets).* Retrieved 2021-05-24, from https://ec.europa.eu/clima/policies/ets_en

European Commission. (2021b). *Kyoto 1st commitment period (2008–12).* Retrieved 2021-05-24, from https://ec.europa.eu/clima/policies/strategies/progress/kyoto_1_en

European Commission. (2021c). *Kyoto 2nd commitment period (2013–20).* Retrieved 2021-05-24, from https://ec.europa.eu/clima/policies/strategies/progress/kyoto_2_en

European Commission. (2021d). *Market stability reserve.* Retrieved 2021-05-24, from https://ec.europa.eu/clima/policies/ets/reform_en

European Environment Agency. (2014). *Eu ets data viewer: manual and user guide.*

European Environment Agency. (2020). *Trends and projections in europe 2020.* doi: 10.2800/830157

Faraway, J. J. (2002). *Practical regression and anova using r.* (Vol. 168). University of Bath Bath.

Flachsland, C., Pahle, M., Burtraw, D., Edenhofer, O., Elkerbout, M., Fischer, C., … Zetterberg, L. (2020). How to avoid history repeating itself: the case for an eu emissions trading system (eu ets) price floor revisited. *Climate Policy*, *20*(1), 133–142.

Gates, B. (2021). *How to avoid a climate disaster*. Dublin.

Gerlagh, R., Heijmans, R. J. R. K., & Rosendahl, K. E. (2020). Covid-19 tests the market stability reserve. *Environmental and Resource Economics*, *76*, 855-865. Retrieved from https://www.sciencedirect.com/science/article/pii/S0095069620301285 doi: https://doi.org/10.1007/s10640-020-00441-0

Gerlagh, R., Heijmans, R. J. R. K., & Rosendahl, K. E. (2021, 03). An Endogenous Emissions Cap Produces a Green Paradox. *Economic Policy*. Retrieved from https://doi.org/10.1093/epolic/eiab011 (eiab011) doi: 10.1093/epolic/eiab011

Gugler, K., Haxhimusa, A., & Liebensteiner, M. (2021). Effectiveness of climate policies: Carbon pricing vs. subsidizing renewables. *Journal of Environmental Economics and Management*, *106*, 102405. Retrieved from https://www.sciencedirect.com/science/article/pii/S0095069620301285 doi: https://doi.org/10.1016/j.jeem.2020.102405

Hintermann, B., Peterson, S., & Rickels, W. (2016). Price and market behavior in phase ii of the eu ets: A review of the literature. *Review of Environmental Economics and Policy*, *10*(1). doi: https://doi.org/10.1093/reep/rev015

Hirst, D. (2018). *Carbon price floor (cpf) and the price support mechanism.*

Hoffmann, V. H. (2007). Eu ets and investment decisions:: The case of the german electricity industry. *European Management Journal*, *25*(6), 464-474. Retrieved from https://www.sciencedirect.com/science/article/pii/S0263237307000837 (Business, Climate Change and Emissions Trading) doi: https://doi.org/10.1016/j.emj.2007.07.008

IEA. (2019). *United kingdom 2019 review.*

IEA. (2021). *Germany 2020 energy policy review.*

IEA. (2021). *The role of china's ets in power sector decabonisation.* Retrieved from https://www.iea.org/reports/the-role-of-chinas-ets-in-power-sector-decarbonisation

International Carbon Action Partnership. (2021). *Korea emissions trading scheme.*

Koch, N., Fuss, S., Grosjean, G., & Edenhofer, O. (2014). Causes of the eu ets price drop: Recession, cdm, renewable policies or a bit of everything?—new evidence. *Energy Policy*, *73*, 676-685. Retrieved from https://www.sciencedirect.com/science/article/pii/S0301421514003966 doi: https://doi.org/10.1016/j.enpol.2014.06.024

Laing, T., Sato, M., Grubb, M., & Comberti, C. (2013). *Assessing the effectiveness of the eu emissions trading scheme.*

Marion, L. (2019). Carbon pricing and power sector decarbonisation: Evidence from the uk. doi: 10.13140/RG.2.2.19205.50400

Metcalf, G. H., & Stock, J. H. (2020). Measuring the macroeconomic impact of carbon taxes. *Harvard Environmental Economics Program*.

Muûls, M., Colmer, J., Martin, R., & Wagner, U. J. (2016). Evaluating the eu emissions trading system: Take it or leave it? an assessment of the data after ten years.

NTV. (2021). *Grüne bleiben beim co2-preis auf linie.* Retrieved 2021-08-12, from https://www.n-tv.de/politik/Gruene-bleiben-beim-CO2-Preis-auf-Linie-article22614257.html

Oestreich, A. M., & Tsiakas, I. (2015). Carbon emissions and stock returns: Evidence from the eu emissions trading scheme. *Journal of Banking & Finance*, *58*, 294-308. Retrieved from https://www.sciencedirect.com/science/article/pii/S0378426615001326 doi: https://doi.org/10.1016/j.jbankfin.2015.05.005

Our World in Data. (2021). *Electricity mix.* Retrieved 2021-08-02, from https://ourworldindata.org/electricity-mix

Perino, G., & Willner, M. (2016). Procrastinating reform: The impact of the market stability reserve on the eu ets. *Journal of Environmental Economics and Management*, *80*, 37-52. Retrieved from https://www.sciencedirect.com/science/article/pii/S0095069616302935 (The economics of the European Union Emission Trading System (EU ETS) market stability reserve) doi: https://doi.org/10.1016/j.jeem.2016.09.006

Phillips, P. C. B., & Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, *67*(5), 1057-1111. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00070 doi: https://doi.org/10.1111/1468-0262.00070

Renewable UK. (2021). *Wind energy.* Retrieved 2021-08-11, from https://www.renewableuk.com/page/WindEnergy

Rogge, K. S., Schneider, M., & Hoffmann, V. H. (2011). The innovation impact of the eu emission trading system — findings of company case studies in the german power sector. *Ecological Economics*, *70*(3), 513-523. Retrieved from https://www.sciencedirect.com/science/article/pii/S0921800910003915 doi: https://doi.org/10.1016/j.ecolecon.2010.09.032

Schmidt, L. (2020). Puncturing the waterbed or the new green paradox? the effectiveness of overlapping policies in the eu ets under perfect foresigth and myopia. *EWI Working Paper*(20/07).

Schwerdt, G., & Woessmann, L. (2020). Empirical methods in the economics of education. *The Economics of Education*, 3–20.

Sijm, J., Neuhoff, K., & Chen, Y. (2006, May). CO2 cost pass through and windfall profits in the power sector [Cambridge Working Papers in Economics]. (0639). Retrieved from https://ideas.repec.org/p/cam/camdae/0639.html

Smale, R., Hartley, M., Hepburn, C., Ward, J., & Grubb, M. (2006). The impact of co2 emissions trading on firm profits and market prices. *Climate Policy*, *6*(1), 31-48. Retrieved from https://doi.org/10.1080/14693062.2006.9685587 doi: 10.1080/14693062.2006.9685587

The United Nations. (2021). *The paris agreement.* Retrieved 2021-08-02, from https://www.un.org/en/climatechange/paris-agreement

UK Government. (2010). *Carbon price floor: support and certainty for low-carbon investment.*

UK Government. (2011). *Budget 2011.*

University of Virginia. (2015). *Understanding q-q plots.* Retrieved 2021-08-09, from https://data.library.virginia.edu/understanding-q-q-plots/

Wilson, G., & Staffell, I. (2018). Rapid fuel switching from coal to natural gas through effective carbon pricing. *Nature Energy*, *3*, 365-372. doi: https://doi.org/10.1038/s41560-018-0109-0

Winchester, N., & Reilly, J. M. (2019). The economic, energy, and emissions impacts of climate policy in south korea. *Climate Change Economics*, *10*(3). doi: 10.1142/S2010007819500106

Wirth, H. (2021). *Recent facts about photovoltaics in germany.* Retrieved from https://www.iea.org/reports/the-role-of-chinas-ets-in-power-sector-decarbonisation

World Nuclear Association. (2021a). *Nuclear power in germany.* Retrieved 2021-06-28, from https://world-nuclear.org/information-library/country-profiles/countries-g-n/germany.aspx

World Nuclear Association. (2021b). *Nuclear power in the united kingdom.* Retrieved 2021-06-28, from https://world-nuclear.org/information-library/country-profiles/countries-t-z/united-kingdom.aspx

Junior Management Science 7(3) (2022) 756-801

Junior Management Science

journal homepage: www.jums.academy

The Impact of Sustainable Supply Chain Management on Corporate Performance – An Empirical Analysis of Manufacturing and Processing Companies in Germany

Der Einfluss von Sustainable Supply Chain Management auf die Unternehmensperformance – Eine empirische Analyse herstellender und verarbeitender Unternehmen in Deutschland

Sören Schwulera

Georg-August-Universität Göttingen

Abstract

Companies implement Sustainable Supply Chain Management (SSCM) practices to remain competitiveness not only on the economic, but also on the environmental and social levels of the Tripple Bottom Line (TBL). The aim of this paper was to empirically investigate the impact of SSCM practices on the economic, environmental, and the social level of corporate performance of manufacturing and processing companies. In order to achieve this goal, a theoretical research model was set up based on relevant literature with four internal and four external SSCM practices, each of them was expected to have a positive effect on all levels of corporate performance. After an online survey of the 500 biggest manufacturing and processing companies in Germany measured by turnover, 61 questionnaires were evaluated using partial least squares structural equation modelling. In total, 10 of the 28 expected positive effects of internal and external SSCM practices on the three levels of corporate performance could be confirmed. This paper provides a theoretical research model for further studies and supports manager in companies in case of implementation of SSCM practices.

Zusammenfassung

Unternehmen implementieren Sustainable Supply Chain Management (SSCM) Praktiken, um nicht nur auf ökonomischer, sondern auch auf ökologischer und sozialer Ebene der Tripple Bottom Line (TBL) wettbewerbsfähig zu bleiben. Das Ziel der vorliegenden Arbeit war es empirisch den Einfluss von SSCM Praktiken auf die ökonomische, ökologische und auch die soziale Ebene der Unternehmensperformance von herstellenden und verarbeitenden Unternehmen zu untersuchen. Zur Zielerreichung wurde anhand einschlägiger Fachliteratur ein theoretisches Forschungsmodell mit jeweils vier internen und externen SSCM Praktiken aufgestellt, für die ein positiver Effekt auf alle drei Ebenen der Unternehmensperformance erwartet wurde. Nach einer Online-Befragung in den 500 umsatzstärksten herstellenden und verarbeitenden Unternehmen in Deutschland wurden 61 Fragebögen anhand einer Partial Least Squares Strukturgleichungsmodellierung ausgewertet. Es konnten insgesamt 10 der 28 erwarteten positiven Effekte interner und externer SSCM Praktiken auf die drei Ebenen der Unternehmensperformance bestätigt werden. Die vorliegende Arbeit bietet ein theoretisches Forschungsmodell für weiterführende Studien und dient Entscheidungsträgern in Unternehmen als Entscheidungshilfe zur Implementierung von SSCM Praktiken.

Keywords: Sustainable Supply Chain Management; Unternehmensperformance; Tripple Bottom Line; Partial Least Squares Strukturgleichungsmodellierung.

1. Einleitung

„The achievement of sustained and equitable development remains the greatest challenge facing the human race“ (World Bank, 1992, S. 1). Die Besorgnisse der Menschen, der Regierungen und der Wirtschaft über die Bedrohungen durch den Klimawandel und die Erschöpfung natürlicher Ressourcen haben in den letzten Jahrzehnten deutlich zugenommen (Vachon & Mao, 2008, S. 1552; Hsu, Choon Tan, Hanim

DOI: https://doi.org/10.5282/jums/v7i3pp756-801

Mohamad Zailani & Jayaraman, 2013, S. 656). Der Klimawandel, dessen Haupttreiber die anthropogene Emission von Treibhausgasen (THG) ist (WMO, 2020, S. 7), ist derzeit eines der größten globalen Probleme (Sames & Köpke, 2012, S. 1; Subramanian & Abdulrahman, 2017, S. 1168). THG Emissionen tragen seit den 1950er Jahren zu einer beispiellosen Erwärmung des Weltklimas bei (IPCC, 2014, S. 2). Infolgedessen befassen sich Regierungen und Unternehmen zunehmend mit Maßnahmen, um diese nachteiligen Auswirkungen auf die Umwelt zu verringern (Huang, Weber & Matthews, 2009, S. 8509; Sharma & Gandhi, 2016, S. 332).

Industrielle Prozesse verursachen einen signifikanten Anteil der globalen THG Emissionen und weisen somit ein hohes Minderungs- und Anpassungspotential auf (Charkovska et al., 2019, S. 907). In Deutschland war der Industriesektor für ca. 20,70 % der gesamten THG Emissionen des Jahres 2016 verantwortlich und ist somit nach dem Energiesektor die zweitgrößte Emissionsquelle (BMU, 2018, S. 34). Besonders die Supply Chains von Industrieunternehmen machen dabei einen großen Anteil der THG Emissionen aus (Sharma & Gandhi, 2016, S. 332). Nach Huang et al. (2009, S. 8509) sind im Durchschnitt mehr als 75 % des CO_2-Fußabdrucks eines Industriesektors auf dessen Supply Chain zurückzuführen. Traditionell beschäftigt sich das Supply Chain Management (SCM) mit den wirtschaftlichen Zielen eines Unternehmens (Das, 2017, S. 1344). Trends wie Globalisierung, Digitalisierung, Outsourcing, Verkürzung von Produktlebenszyklen und die hohen Erwartungen der Kunden an Preis und Qualität, aber auch Ressourcenknappheit und strenge Umweltvorschriften führen zu einem immer komplexer werdenden Aufgabenspektrum für das SCM, das über die traditionellen Aufgaben hinausgeht (Saeed & Kersten, 2019, S. 1; Ageron, Gunasekaran & Spalanzani, 2012, S. 168). Langfristige und enge Beziehungen zu Lieferanten, Kunden sowie anderen strategischen Partnern entlang der Supply Chain sind zu wichtigen Schlüsselfaktoren für die Wettbewerbsfähigkeit von Unternehmen geworden, wobei seit einiger Zeit nicht nur monetäre Ziele im Vordergrund stehen (Andersen & Skjoett-Larsen, 2009, S. 75; Silva, Guarnieri, Carvalho, Farias & Reis, 2019, S. 3).

Diese Entwicklungen der letzten Jahre zeigen, dass die bloße Verfolgung wirtschaftlicher Ziele aus Sicht langfristiger Rentabilität und Nachhaltigkeit keine solide Entscheidungsgrundlage mehr darstellt (Das, 2017, S. 1344). Mit steigendem Interesse an Umwelt- und Sozialthemen üben interne und externe Stakeholder, wie z.B. das Management, Mitarbeiter, Aktionäre, Kunden, Lieferanten oder Regierungen immer mehr Druck auf Unternehmen und deren Supply Chains aus (Andersen & Skjoett-Larsen, 2009, S. 75; Seuring, Sarkis, Müller & Rao, 2008, S. 1545). Weisen unternehmerische Aktivitäten in ökologischer oder sozialer Hinsicht Defizite auf, kann dies nicht nur erhebliche Auswirkungen auf den Umsatz oder die Bewertung des Unternehmens haben, sondern im schlimmsten Fall auch zu Strafzahlungen führen. Diese Umstände haben viele Unternehmen dazu motiviert, sogenannte Green Supply Chain Management (GSCM) Praktiken zu integrieren, um die negativen Auswirkungen ihrer Prozesse und Produkte auf die Umwelt zu begrenzen (Vachon & Klassen, 2008, S. 299; Das, 2018, S. 5776). Gleichzeitig führen Unternehmen seit den letzten Jahren Corporate Social Responsibility (CSR) Praktiken in ihre Geschäftsabläufe ein, um auch auf die sozialen Bedürfnisse interner und externer Stakeholder einzugehen (Das, 2017, S. 1344). Die simultane Berücksichtigung der ökonomischen, ökologischen und sozialen Dimensionen im SCM entsprechen dem Konzept der Triple Bottom Line (TBL) nach (Elkington, 1999, S. 397). Sustainable Supply Chain Management (SSCM) vereint die Ziele von GSCM und CSR, was dem Unternehmen letztendlich bei der Erreichung der ökonomischen, ökologischen und sozialen Ziele helfen soll (Das, 2017, S. 1344). Dies zeigt, dass ein erfolgreiches SSCM sowie die Performance der Supply Chain zu wichtigen Faktoren für die Wettbewerbsfähigkeit und die Performance von Unternehmen jeglicher Größenordnungen und Branchenzugehörigkeiten geworden sind (Seuring et al., 2008, S. 1545; Seuring, 2013, S. 1513).

Das akademische und unternehmerische Interesse an Themen zum SSCM nimmt deutlich zu (Seuring & Müller, 2008, S. 1699). In den letzten Jahren wurde eine Reihe an wissenschaftlichen Beiträgen zum Zusammenhang zwischen SSCM und der Unternehmensperformance publiziert. In vielen empirischen Studien wird der Einfluss von GSCM Praktiken auf die ökonomische und ökologische Performance von Unternehmen untersucht. Die soziale Dimension wird in diesen Publikationen jedoch häufig vernachlässigt (Das, 2017, S. 1344; Panigrahi, Bahinipati & Jain, 2019, S. 1027). Die meisten Autor*innen, die in ihren Publikationen neben der ökologischen und ökonomischen auch die soziale Dimension betrachten, entwickelten theoretische Konzepte auf Basis systematischer Literaturreviews (Das, 2018, S. 5776). Nur wenige empirische Studien befassen sich mit dem Einfluss von SSCM Praktiken auf die gesamte Unternehmensperformance nach der TBL (Das, 2017, S. 1345). Aufgrund der Aktualität dieser Thematik und um die beschriebene Forschungslücke zu behandeln, soll im Rahmen der vorliegenden quantitativ empirischen Arbeit anhand einer Primärforschung der Einfluss von SSCM auf die Unternehmensperformance im Sinne des TBL-Konzepts untersucht werden. Zur Erreichung dieses Ziels sollen folgende Forschungsfragen beantwortet werden:

FF1: *Wie lässt sich das Konzept des SSCM definieren?*

FF2: *Welche SSCM Praktiken können aus der aktuellen Fachliteratur identifiziert werden und wie lassen sie sich klassifizieren?*

FF3: *Wie und in welchem Maße wirken sich die identifizierten SSCM Praktiken auf die ökologische, ökonomische und soziale Performance herstellender und verarbeitender Unternehmen in Deutschland aus?*

Mit der empirischen Untersuchung soll ein wichtiger Beitrag für die Theorie und Praxis geleistet werden. Für die SSCM-Forschung soll die vorliegende Arbeit ein theoretisches Forschungsmodell bieten, welches nicht nur die ökologische und ökonomische, sondern auch die soziale Dimension der

TBL beinhaltet. Dieses Modell soll somit als Grundlage für weitere Studien dienen. Für die Praxis sollen unternehmerische Entscheidungsträger mithilfe der vorliegenden Arbeit eine Entscheidungshilfe zur Implementierung verschiedener etablierter SSCM Praktiken aus einschlägiger Literatur bekommen. Um dies zu erreichen, wird eine Online-Befragung nach der Survey Research Methodik in herstellenden und verarbeitenden Unternehmen in Deutschland durchgeführt. Die gewonnenen Primärdaten werden dann mithilfe des Partial Least Squares (PLS)-Ansatzes analysiert.

Im Detail wird zur Erreichung der Zielsetzung und zur Beantwortung der Forschungsfragen folgendermaßen vorgegangen: Im sich anschließenden Kapitel 2 werden die wesentlichen theoretischen Grundlagen gegeben. Hier wird die Begrifflichkeit des SSCM anhand des Nachhaltigkeitsbegriffs und des SCM hergeleitet sowie eine für diese Arbeit relevante Definition festgelegt. Anschließend werden Treiber und Praktiken des SSCM anhand einschlägiger Literatur identifiziert und ein sachlogischer Zusammenhang der SSCM Praktiken zur Unternehmensperformance hergestellt, indem ein theoretisches Hypothesenmodell aufgestellt wird, welches die Grundlage für die empirische Untersuchung bildet. Das dritte Kapitel gibt einen Überblick über die Methodik von Survey Research, das Verfahren zur Datenerhebung sowie die zur Datenanalyse genutzte Methodik der PLS-Strukturgleichungsmodellierung (PLS-SGM). Weiterhin werden in diesem Kapitel die wichtigsten Charakteristika der zugrundeliegenden Stichprobe vorgestellt. Im darauffolgenden vierten Kapitel werden die Ergebnisse anhand verschiedener, für die PLS-SGM typischen Gütekriterien auf Reliabilität und Validität geprüft, bevor anschließend die Ergebnisse hinsichtlich der hergeleiteten Hypothesen systematisch präsentiert werden. Im fünften Kapitel werden die Ergebnisse der Befragung diskutiert und Implikationen für die Praxis gegeben. Weiterhin werden Limitationen der vorliegenden Arbeit aufgezeigt und das hieraus resultierende Potenzial für weitere Forschung erläutert. Abschließend werden im sechsten Kapitel die wichtigsten Ergebnisse hinsichtlich der übergeordneten Zielsetzung und zur Beantwortung der Forschungsfragen zusammengefasst und die vorliegende Arbeit in den aktuellen Forschungsstand eingeordnet.

2. Theoretische Grundlagen

Das Kapitel der theoretischen Grundlagen gibt einen Überblick über die relevanten Rahmenaspekte der vorliegenden Arbeit in Bezug auf die Problemstellung. Zu Beginn wird das Konzept der Nachhaltigkeit erläutert. Dazu wird im ersten Teil von Kapitel 2.1 auf die Entwicklung des Nachhaltigkeitsbegriffs eingegangen und im zweiten Teil auf Nachhaltigkeit in Unternehmen. In Kapitel 2.2 wird eine Definition des SCM vorgenommen und dessen Aufgaben, Ziele und Konzepte vorgestellt. Das darauffolgende Kapitel 2.3 dient der Verknüpfung des Nachhaltigkeitsbegriffs und SCM zum SSCM. Abschließend wird in Kapitel 2.4 der Zusammenhang zwischen SSCM Praktiken und der Unternehmensperformance im Sinne der TBL hergestellt und anhand von Hypothesen in einem theoretischen Forschungsmodell zusammengefasst.

2.1. Konzept der Nachhaltigkeit

Bevor auf das Konzept des SSCM eingegangen werden kann, ist es wichtig, zunächst eine für diese Arbeit relevante Definition des Nachhaltigkeitsbegriffs festzulegen. Dazu wird im Folgenden die Entwicklung des Nachhaltigkeitsbegriffs dargestellt. Im Anschluss wird auf das Nachhaltigkeitsmanagement in Unternehmen eingegangen.

2.1.1. Entwicklung des Nachhaltigkeitsbegriffs

Der Nachhaltigkeitsbegriff ist aufgrund unterschiedlicher Perspektiven und Ansichten sehr umstritten. Viele Akteure, z.B. in Politik, Wirtschaft oder Umweltorganisationen, interpretieren und vertreten nachhaltige Entwicklung auf ihre eigene Weise (Giddings, Hopwood & O'Brien, 2002, S. 187). In diesem Teilkapitel wird deshalb einerseits eine definitorische Grundlage des Nachhaltigkeitsbegriffs für die vorliegende Arbeit geschaffen, andererseits werden die bedeutendsten Meilensteine des Begriff-Ursprungs beschrieben, die zu der gegenwärtigen Bedeutsamkeit des Begriffs beigetragen haben.

Seinen Ursprung nimmt der Nachhaltigkeitsbegriff bereits Anfang des 18. Jahrhunderts im Bereich der Forstwirtschaft (Bretzke & Barkawi, 2012, S. 13; Altundas, Memeti, Rau & Schrag, 2015, S. 16; Brüssel, 2018, S. 12). Der kursächsische Oberberghauptmann Hans Carl von Carlowitz (1645-1714) gilt mit seinem Werk „Sylvicultura oeconomica" von 1713 als Begründer des Nachhaltigkeitsprinzips. Holz war wichtigster Rohstoff, der vor allem zum Bauen von Häusern und Bergstollen, den Abbau und das Schmelzen von Erz, zum Kochen und Heizen sowie für vorindustrielle Produktionsprozesse wie den Schiffbau genutzt wurde (Colsman, 2013, S. 11–12). Aufgrund dieser exzessiven Ausnutzung der Waldflächen kam es zu einer Holzknappheit in Deutschland (von Carlowitz, Thomasius & Bendix, 2013, S. 47–50). Zur Lösung dieser Problematik schlug von Carlowitz vor, die Abholzung von Waldflächen nur bis zu dem Maße zu gestatten, wie durch Aufforstung wieder nachwachsen kann (Müller-Christ, 2010, S. 104–105; von Hauff, 2014, S. 1), um die langfristige Verfügbarkeit des Rohstoffes zu gewährleisten (Dahm, 2019, S. 120; Carnau, 2011, S. 12).

Grundlegend für die heutige Nachhaltigkeitsdebatte ist die Veröffentlichung der 1972 vorgestellten Studie „The Limits to Growth" des Club of Rome (Bretzke & Barkawi, 2012, S. 12; von Hauff, 2014, S. 1; Gogoll & Wenke, 2017, S. 120). Meadows, Meadows, Randers und Behrens (1972) machten in ihrem Bericht darauf aufmerksam, dass „eine Fortschreibung der aktuellen Trends hinsichtlich des Bevölkerungswachstums und der Nachfrage nach nichtregenerativen Ressourcen bis Mitte des 21. Jahrhunderts zu einer großen wirtschaftlichen Beeinträchtigung führen würde" (von Hauff, 2014, S. 6). Aus heutiger Sicht gelten die Prognosen der Studie zwar als zu pessimistisch, dennoch wurde dadurch eine noch nie dagewesene gesellschaftliche und politische Diskussion über mögliche Auswirkungen von

Ressourcenknappheit auf das Wirtschaftswachstum und die Weltbevölkerung angeregt (Gogoll & Wenke, 2017, S. 120; Müller-Christ, 2010, S. 33). Noch im selben Jahr fand die erste Umweltkonferenz der UN (United Nations Conference on the Human Environment, UNCHE) in Stockholm statt, welche als Beginn der internationalen Zusammenarbeit zum Thema der nachhaltigen Entwicklung gilt (Boone, Jayaraman & Ganeshan, 2012, S. 4; von Hauff, 2014, S. 1). Aus der Umweltkonferenz ging eine Erklärung über die menschliche Umwelt sowie ein Handlungskonzept zum Schutz und Verbesserung der Umwelt hervor (Müller-Christ, 2010, S. 35). Auf der UNCHE wurde zudem das „United Nations Environmental Program“ (UNEP) gegründet, was dazu führte, dass in vielen teilnehmenden Staaten der Konferenz Umweltministerien eingerichtet wurden. Letztendlich gilt die UNCHE aber nur als Auftakt zwischenstaatlicher Zusammenarbeit zum Thema Nachhaltigkeit (Boone et al., 2012, S. 4). Im Jahr 1979 fand die erste Weltklimakonferenz der World Meteorological Organization (WMO) in Genf statt, bei welcher Experten aus der Klimaforschung über den globalen Klimawandel diskutierten (WMO, 1979, S. VIII; Bretzke & Barkawi, 2012, S. 12).

Der 1987 von der World Comission on Evironment and Development (WCED) veröffentlichte Brundtland-Bericht „Our Common Future“ ist ein weiterer Meilenstein in der internationalen Nachhaltigkeitsdebatte (WCED, 1987; Bretzke & Barkawi, 2012, S. 12; Mann & Kaur, 2020, S. 60). Die WCED war 1984 durch eine Initiative der UNEP eingesetzt worden, nachdem die UN 1982 feststellen musste, dass die 1972 in Stockholm festgelegten Ziele der UNCHE nicht zu erreichen waren. Folglich war die WCED vor dem Hintergrund wachsender ökologischer, ökonomischer und sozialer Probleme dazu berufen worden, kritische Fragen zum Zusammenhang von Umweltschutz und wirtschaftlicher Entwicklung noch einmal zu prüfen und konkrete Aktionsvorschläge zu formulieren. Außerdem sollten Konzepte zur Stärkung der internationalen Zusammenarbeit im Bereich Nachhaltigkeit ausgearbeitet werden (WCED, 1987, S. 241; Müller-Christ, 2010, S. 35; Albino, 2013, S. 4; von Hauff, 2014, S. 8). Der Abschlussbericht der WCED, der nach der Kommissionsvorsitzenden Gro Harlem Brundtland benannt wurde, prägte die wohl am meisten zitierte Definition von Nachhaltigkeit bzw. nachhaltiger Entwicklung (z. B. Linton, Klassen & Jayaraman, 2007, S. 1076; Carter & Rogers, 2008, S. 363; Vachon & Mao, 2008, S. 1553; Z. Wang & Sarkis, 2013, S. 873; Mann & Kaur, 2020, S. 60). Die Brundtland-Definition lautet wörtlich:

„Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs. It contains within it two key concepts:

- the concept of ‘needs’, in particular the essential needs of the world’s poor, to which overriding priority should be given; and
- the idea of limitations imposed by the state of technology and social organization on the environment’s ability to meet present and future needs” (WCED, 1987, S. 37).

In dieser Definition betont die WCED besonders den Aspekt der Langfristigkeit des Nachhaltigkeitskonzepts und das Prinzip der Gerechtigkeit zwischen der gegenwärtigen und zukünftigen Generation (Diesendorf, 2000, S. 21). Weiterhin wird die enge Verknüpfung zwischen Armutsbekämpfung, Verbesserung der Umwelt und sozialer Gleichberechtigung durch nachhaltiges Wachstum der Wirtschaft hervorgehoben. Aufgrund des hohen Interpretationsspielraums hat diese Definition ihre breite Akzeptanz gefunden (Mebratu, 1998, S. 501–502). In der Literatur wird die Brundtland-Definition jedoch aufgrund mangelnder Klarheit bezüglich ihrer Anwendbarkeit und des geringen Konkretisierungsgrads aber auch häufig kritisiert (Vachon & Mao, 2008, S. 1553). Costanza und Patten (1995, S. 193) weisen diese Kritik zurück, denn „das Konzept der Nachhaltigkeit sei weniger eine exakte Definition von Nachhaltigkeit, sondern es gehe bei der Bestimmung von nachhaltiger Entwicklung vielmehr um die Bestimmung dessen, was Bestand haben soll […] sowie darum, auf die miteinander verknüpften zeitlichen und räumlichen Ebenen zu verweisen und deutlich zu machen, dass eine Politik der Nachhaltigkeit diese mit einzubeziehen hat“ (Petschow, Droge, Hübner & Meyerhoff, 1997, S. 115). Ferner werden bei Anwendung der Brundtland-Definition in der Literatur die Begriffe Nachhaltigkeit und nachhaltige Entwicklung häufig synonym verwendet (Carnau, 2011, S. 13). Streng genommen handelt es sich bei dem Begriff Nachhaltigkeit um ein Ziel, was erreicht werden soll, und bei nachhaltiger Entwicklung um den Prozess, um jenes Ziel zu erreichen (Sikdar, 2003, S. 1928–1929). Da ökologische, ökonomische und soziale Systeme aber einer ständigen Veränderung ausgesetzt sind, wird davon ausgegangen, dass Nachhaltigkeit keinen endgültigen Zustand darstellt, sondern einen Prozess von stetiger Anpassung (Gallopín, 2003, S. 19), weshalb die Begriffe gleichgesetzt werden können. Initiiert durch die Brundtland-Kommission fand 1992 in Rio de Janeiro die United Nations Conference on Environment and Development (UNCED), der sogenannte Erdgipfel, statt (Albino, 2013, S. 4). Durch Unterzeichnung der bei der UNCED entwickelten Klimarahmenkonvention der Vereinten Nationen (UNFCCC) verpflichten sich aktuell 196 Staaten sowie die gesamte EU zum Handeln gegen den Klimawandel (BMU, 2018, S. 19).

Im Laufe der Nachhaltigkeitsdebatte haben sich drei klassische Dimensionen der Nachhaltigkeit etabliert: die ökologische, ökonomische und soziale Dimension (Bretzke & Barkawi, 2012, S. 28; Colsman, 2013, S. 14). Die ökologische Dimension bezieht sich auf den Schutz der Umwelt und Natur für die kommenden Generationen durch z. B. weniger Emissionen und Abfälle sowie geringeren Energie- und Ressourcenverbrauch. Die ökonomische Dimension fokussiert die Schaffung von dauerhaftem Wohlstand und finanziellem Wertzuwachs. Hinsichtlich der sozialen Dimension steht eine lebenswerte und zukunftsfähige Gesellschaft im Mittelpunkt (Altundas et al., 2015, S. 17; Colsman, 2013, S. 14). Dieses

sogenannte Drei-Säulen-Modell der Nachhaltigkeit oder auch TBL, beruht auf „Canibals with forks“ von Elkington (1999, S. 397). Der Autor definiert die TBL als nachhaltige Entwicklung, welche das simultane Streben nach ökologischer Umweltqualität, ökonomischen Wohlstand und sozialer Gerechtigkeit beinhaltet. Dabei bezieht sich der Autor explizit auf Unternehmen, die sich nicht nur auf ihre finanziellen Ziele, sondern auf die Ziele aller drei Dimensionen gleichermaßen konzentrieren sollten (Elkington, 1999, S. 397; Das, 2017, S. 1346).

2.1.2. Nachhaltigkeit in Unternehmen

Ohne den Beitrag von Unternehmen ist eine nachhaltige Entwicklung nicht möglich, da diese einen großen Einfluss auf ökologische und soziale Interessen haben (Schaltegger, Lüdeke-Freund & Hansen, 2012, S. 96). Unternehmen entscheiden über ihren Verbrauch an Ressourcen, die Menge an Emissionen aber auch über Arbeitsplätze und -bedingungen (Kurz & Wild, 2015, S. 323). Wird die Brundtland-Definition auf Unternehmensebene übertragen, kann Nachhaltigkeit in Unternehmen entsprechend so definiert werden, dass die Bedürfnisse direkter und indirekter Stakeholder (z. B. Mitarbeiter, Aktionäre, Gesellschaft etc.) erfüllt werden, ohne dass die Erfüllung der Bedürfnisse zukünftiger Stakeholder beeinträchtigt wird (Dyllick & Hockerts, 2002, S. 131). Nach Shrivastava (1995, S. 955) ist Nachhaltigkeit in Unternehmen „the potential for reducing long-term risks associated with resource depletion, fluctuations in energy costs, product liabilities, and pollution and waste management.“ Hier wird Nachhaltigkeit aber lediglich aus ökologischer Sicht betrachtet. Demzufolge wird im Folgenden die Definition von Sikdar (2003, S. 1928) verwendet. So ist unternehmerische Nachhaltigkeit „a wise balance among economic development, environmental stewardship, and societal equity” (Sikdar, 2003, S. 1928), was mit der TBL nach Elkington (1999, S. 397) einhergeht.

Ein ökonomisch nachhaltiges Unternehmen sollte nach Dyllick und Hockerts (2002, S. 133) stets einen zufriedenstellenden Cashflow garantieren, um nicht nur die Liquidität zu sichern, sondern auch, um eine langfristig überdurchschnittliche Rendite für seine Shareholder zu erzielen. Ein ökologisch nachhaltiges Unternehmen nutzt ausschließlich natürliche Ressourcen und verursacht keine Emissionen, die nicht auf natürliche Weise absorbiert werden können. Es werden also keinerlei Aktivitäten ausgeübt, die die Leistung des Ökosystems beeinträchtigen. Ein sozial nachhaltiges Unternehmen sollte einen Mehrwert für die Gesellschaft schaffen, indem es das Humankapital erhöht und das gesellschaftliche Kapital, wie z. B. Investitionen in Bildung und Kultur, fördert (Dyllick & Hockerts, 2002, S. 133–134).

Während das Management von vielen Unternehmen noch vor einigen Jahren eine Nachhaltigkeitsstrategie als unnötigen Kostenfaktor angesehen hat, haben Führungskräfte nun erkannt, dass die Beachtung von Nachhaltigkeitsprinzipien als Chance für langfristige Wettbewerbsvorteile und die Zukunft des Unternehmens dienen kann (Millar, Hind & Magala, 2012, S. 490). Die Integration eines Nachhaltigkeitsmanagements dient dem Ziel der unternehmerischen Nachhaltigkeit, indem das Unternehmen nicht nur eine integrative Berücksichtigung ökologischer, ökonomischer und sozialer Ziele anstrebt, sondern auch zur nachhaltigen Entwicklung von Wirtschaft und Gesellschaft beiträgt (Schaltegger, Herzig, Kleiber, Klinke & Müller, 2007, S. 3). Mit dem Ziel nachhaltiger Entwicklung ergeben sich für Unternehmen vier Nachhaltigkeitsherausforderungen (vgl. Abbildung 1).

Die erste Herausforderung (1) ist die Steigerung der Öko-Effektivität bzw. die ökologische Herausforderung (Schaltegger et al., 2007, S. 14). Öko-Effektivität „misst den Grad der absoluten Umweltverträglichkeit, das heißt, wie gut das angestrebte Ziel der Minimierung von Umwelteinwirkung erreicht wurde“ (Schaltegger et al., 2007, S. 15). Gemessen wird die Öko-Effektivität an Kennzahlen, wie z. B. Materialeinsatz pro Wertschöpfung oder den gesamten Energieverbrauch im Produktlebenszyklus (Material- und Energieflussrechnung) (Stahlmann & Clausen, 1999, S. 20). Die zweite Herausforderung (2) stellt die Steigerung der Sozio-Effektivität bzw. die soziale Herausforderung dar. Durch „die Reduktion sozial unerwünschter Auswirkungen des Unternehmens und die Förderung positiver sozialer Wirkungen“ soll die Sozio-Effektivität gesteigert werden (Schaltegger et al., 2007, S. 11). Die dritte Herausforderung (3) ist die ökonomische Herausforderung an das Umwelt- und Sozialmanagement eines Unternehmens, also die Verbesserung von Öko- und Sozio-Effizienz[1]. Schaltegger et al. (2007, S. 17) definieren die Öko-Effizienz als das „Verhältnis zwischen ökonomischen, monetären und einer physikalischen (ökologischen) Größe“ und ist die Kurzform für ökonomisch-ökologische Effizienz. Bei der Messung der Öko-Effizienz fließt eine ökonomische Größe als Wertschöpfung und eine ökologische Größe als Schadschöpfung in das Verhältnis ein (Kicherer, Schaltegger, Tschochohei & Pozo, 2007, S. 537). Die Schadschöpfung wird als „die Summe aller, während eines Produktlebens durch betriebliche Leistungsprozesse direkt und indirekt [...] verursachten Umweltbelastungen“ definiert (Schaltegger & Sturm, 1990, S. 280). Als Kennzahlen der Öko-Effizienz können z. B. der Abfallwirtschaftskostenanteil oder der spezifische Energieverbrauch pro Beschäftigten genannt werden (Stahlmann & Clausen, 1999, S. 20). Die Sozio-Effizienz, auch ökonomisch-soziale Effizienz genannt, ist analog zur Öko-Effizienz das Verhältnis von Wertschöpfung zum sozialen Schaden. Der soziale Schaden ist dabei die Summe der negativen Auswirkungen, die von den Produktionsprozessen eines Unternehmens ausgehen. Als Kennzahl kann z. B. das Verhältnis von Wertschöpfung zu den Krankheitstagen der Beschäftigten genannt werden (Schaltegger et al., 2007, S. 17).

Es muss erwähnt werden, dass eine Steigerung der Öko- und Sozio-Effizienz nicht zwangsläufig mit einer Verbesse-

[1] **Abgrenzung von Effektivität und Effizienz**: Eine Handlung ist effektiv, „wenn sie zweckmäßig ist, d. h. eine bezweckte Zustandsveränderung verursacht.“ Eine Handlung ist hingegen effizient, „wenn die verursachte Zustandsveränderung ohne Verschwendung geschieht, d. h. eine weitergehende Zustandsverbesserung nur bei anderweitiger Verschlechterung möglich wäre“ (Dyckhoff & Ahn, 2001, S. 112).

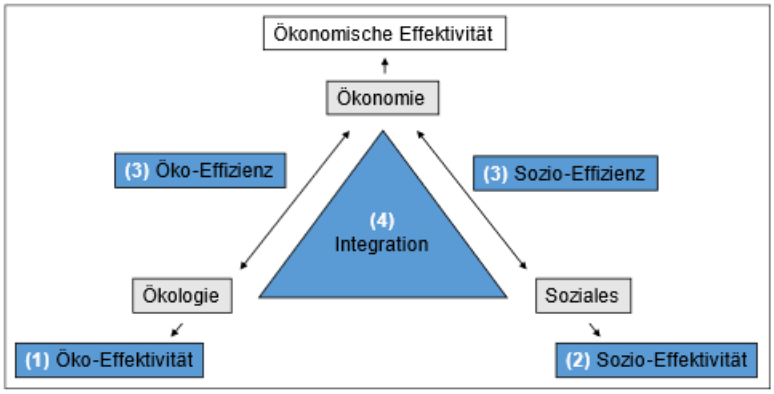


Abbildung 1: Nachhaltigkeitsherausforderungen an Unternehmen

Quelle: eigene Darstellung, nach (Schaltegger et al., 2007, S. 14)

rung der Öko- oder Sozio-Effektivität einhergeht. Im Extremfall kann sich die ökologische oder soziale Situation eines Unternehmens mit einer Verbesserung des Verhältnisses zwischen Wertschöpfung und Schadschöpfung sogar verschlechtern (Schaltegger et al., 2007, S. 17). Dieses Phänomen wird auch als Rebound-Effekt bezeichnet. Wird z. B. durch eine neue Technologie der Produktionsprozess effizienter, können Produkte mit weniger Ressourcen und oft auch zu geringeren Kosten hergestellt werden. Dies kann wiederum zu einem erhöhten Konsumverhalten und Verbrauch dieser Produkte führen (Berkhout, Muskens & Velthuijsen, 2000, S. 426). Aus diesem Grund stehen Unternehmen vor der vierten Herausforderung (4), der sogenannten Integrationsherausforderung. Diese beinhaltet zum einen die Zusammenführung der drei zuvor genannten Herausforderungen und zum anderen die Einbettung des Umwelt- und Sozialmanagements in das ökonomische Management des Unternehmens. Somit soll ein umfassendes Nachhaltigkeitsmanagement gebildet werden, das sowohl zur Steigerung der Öko- und Sozio-Effizienz als auch zur Verbesserung der Öko- und Sozio-Effektivität führt (Schaltegger et al., 2007, S. 14–18). Bevor die Begriffe Nachhaltigkeit, Nachhaltigkeitsmanagement und SCM miteinander verknüpft werden, ist es sinnvoll, im folgenden Teilkapitel vorerst eine Definition des SCM vorzunehmen.

2.2. Supply Chain Management

Einer der bedeutendsten Paradigmenwechsel in der modernen Unternehmensführung der letzten Jahrzehnte ist, dass sich Unternehmen von nun an mittels ihrer Lieferkette unterscheiden und konkurrieren, anstatt als autonome Einheiten zu agieren (Lambert, Cooper & Pagh, 1998, S. 1). Seit den 1990er Jahren wird das Konzept des SCM sowohl in der Theorie als auch in der Praxis diskutiert (Ellram, 1990, S. 1). Da die Wurzeln des SCM in der Unternehmenspraxis liegen, hat sich bislang aber noch kein einheitliches Verständnis in Theorie und Praxis durchsetzen können (Werner, 2017, S. 6). Aufgrund dessen werden im Folgenden Definitionen aus der Literatur diskutiert und eine für diese Arbeit relevante Definition festgelegt. Im zweiten Abschnitt dieses Kapitels wird auf die Aufgaben, Ziele und Konzepte des SCM eingegangen.

2.2.1. Definition des Supply Chain Managements

Im engeren Sinne wird die Supply Chain als Lieferkette oder auch als unternehmensübergreifende Wertschöpfungskette verstanden (Busch & Dangelmaier, 2004, S. 4). Nach Zijm, Klumpp, Heragu und Regattieri (2019, S. 33) umfasst eine Supply Chain alle Aktivitäten, die nötig sind, um Rohstoffe zu Endprodukten zu verarbeiten. Dies beinhaltet neben der Beschaffung, der Herstellung von verschiedenen Komponenten, der Endmontage und den Vertrieb der Endprodukte auch alle erforderlichen Tätigkeiten zum Transport und Lagerung der Materialien (Zijm et al., 2019, S. 33). Diese Definition betont vor allem den Materialfluss in der Supply Chain. Handfield und Nichols (1999, S. 2) erwähnen zusätzlich den interorganisationalen Informationsfluss. Demzufolge enthält eine Supply Chain „all activities associated with the flow and transformation of goods from the raw materials stage (extraction), through the end user, as well as the associated information flows" (Handfield & Nichols, 1999, S. 2). Mentzer et al. (2001, S. 4) gehen noch einen Schritt weiter, indem die Autor*innen zusätzlich den Fluss von Dienstleistungen und finanziellen Mitteln zwischen den beteiligten Unternehmen hervorheben. Die Autor*innen bezeichnen die Supply Chain dementsprechend als „set of three or more entities (organizations or individuals) directly involved in the upstream and downstream flows of products, services, finances, and / or information from a source to a customer" (Mentzer et al., 2001, S. 4). Immer häufiger umfassen Supply Chain Aktivitäten auch die Rücknahme von Produkten sowie das Recycling von Bestandteilen und Materialien (Schaltegger & Harms, 2010, S. 7). Für den weiteren Verlauf der Arbeit wird für die Supply Chain die Definition von Mentzer et al. (2001, S. 4) herangezogen.

In der Unternehmenspraxis handelt es sich bei einer Supply Chain vielmehr um ein Netzwerk von Organisationen, als um eine Kette, da üblicherweise mehrere Lieferanten und Kunden in einem Gesamtsystem miteinander verflochten sind (Lambert et al., 1998, S. 1; Christopher, 2008, S. 5). Dem-

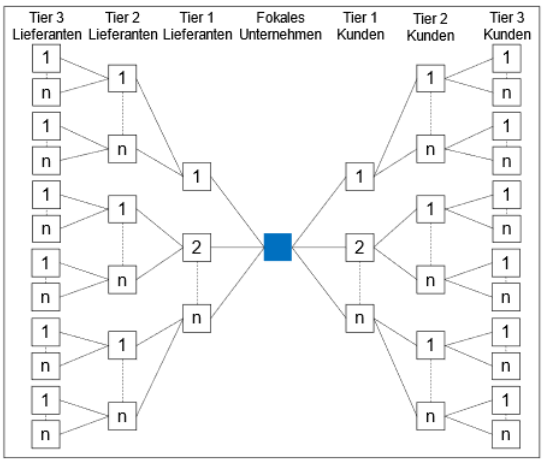


Abbildung 2: Supply Chain Netzwerkstruktur

Quelle: eigene Darstellung, nach (Lambert et al., 1998, S. 3)

nach sollte eine Supply Chain eher als ein Netzwerk voneinander abhängiger Unternehmen bezeichnet werden (Aitken, 1998, S. 1–2). Nach Sydow und Windeler (2001, S. 133) ist ein Unternehmensnetzwerk eine von einem oder mehreren Unternehmungen geführte Organisationsform zur gemeinsamen Generierung von Wettbewerbsvorteilen. Diese Organisationsform zeichnet sich „durch komplex-reziproke, eher kooperative [...] und relativ stabile Beziehungen zwischen rechtlich selbständigen, wirtschaftlich jedoch zumeist abhängigen Unternehmen“ aus (Sydow, 1999, S. 82). Das fokale Unternehmen stellt dabei den Mittelpunkt des Supply Chain Netzwerkes dar (vgl. Abbildung 1). Nach Seuring und Müller (2009, S. 166) nehmen fokale Unternehmen die Führungsrolle in einer Supply Chain ein, indem sie den Marktzugang sicherstellen und für die Endkunden sichtbar sind, die Produkte maßgebend gestalten, dessen grundlegenden Charakteristika festlegen sowie die Lieferanten auswählen und entscheiden, über welche Stufen das Endprodukt zum Kunden gelangt. Wie in Abbildung 2 dargestellt, befinden sich auf den vorgelagerten (upstream) Stufen des fokalen Unternehmens die direkten Lieferanten (Tier 1 Lieferanten), Vorlieferanten (Tier 2 Lieferanten) und weitere Vorlieferanten (Tier 3+n Lieferanten). Analog befinden sich auf den nachgelagerten (downstream) Stufen die direkten Tier 1 Kunden (z. B. Großhandel oder Direktvertrieb), Tier 2 Kunden (z. B. Einzelhandel) und Tier 3+n Kunden (z. B. Endverbraucher) (Lambert et al., 1998, S. 3; Schaltegger & Harms, 2010, S. 7).

In der Literatur existiert eine Vielzahl an Definitionen des SCM (Busch & Dangelmaier, 2004, S. 5). Nach Cooper, Lambert und Pagh (1997, S. 2) ist das SCM die Integration von Geschäftsprozessen entlang der gesamten Supply Chain. Auch bei Mentzer et al. (2001, S. 18) stehen die Geschäftsprozesse im Mittelpunkt der Definition. Sie definieren das SCM als strategische Koordination traditioneller Geschäftsprozesse des fokalen Unternehmens und die der anderen Unternehmen entlang der Supply Chain, um langfristig die Performance aller Beteiligten entlang der gesamten Supply Chain zu verbessern (Mentzer et al., 2001, S. 18). Ähnlich definieren auch Handfield und Nichols (1999, S. 2) das SCM als „integration of these activities[2] through improved supply chain relations, to achieve a sustainable competitive advantage.“ Die Autor*innen ergänzen dem also noch den Material- und Informationsfluss (Handfield & Nichols, 1999, S. 2). Zusammenfassend setzt sich das SCM zum einen aus dem Management von Material- und Informationsfluss, zum anderen aus dem Management der interorganisationalen Zusammenarbeit mit Lieferanten und Kunden zusammen (Seuring & Müller, 2009, S. 166).

Eine Definition, die sowohl den Material- und Informationsfluss als auch die Kooperation entlang des gesamten Supply Chain Netzwerkes beinhaltet, bietet Hahn (1999, S. 851). Demzufolge beinhaltet das SCM die „Planung, Steuerung und Kontrolle des gesamten Material- und Dienstleistungsflusses, einschließlich der damit verbundenen Informations- und Geldflüsse, innerhalb eines Netzwerkes von Unternehmungen, die im Rahmen von aufeinanderfolgenden Stufen der Wertschöpfungskette an der Entwicklung, Erstellung und Verwertung von Sachgütern und / oder Dienstleistungen partnerschaftlich zusammenarbeiten mit dem Ziel der Ergebnis- und Liquiditätsoptimierung - unter Beachtung von sozio-ökologischen Zielen“ (Hahn, 1999, S. 851). Dieser Definition des SCM wird im weiteren Verlauf der Arbeit gefolgt, da sie zum einen mit der festgelegten Definition einer Supply Chain von Mentzer et al. (2001, S. 4) einhergeht und zum anderen nicht nur auf ökonomische Ziele, sondern auch auf sozio-ökologische Ziele eingeht. In nachfolgender Tabelle 1 werden die für die vorliegende Arbeit festgelegten Definitionen von Supply Chain und SCM noch einmal dargestellt.

[2] „these activities” bezieht sich auf die bereits auf S. 10 zitierte Definition der Supply Chain von Handfield und Nichols (1999, S. 2).

Tabelle 1: Definition von Supply Chain und SCM

Definition		Autor*innen
Supply Chain	„Set of three or more entities (organizations or individuals) directly involved in the upstream and downstream flows of products, services, finances, and / or information from a source to a customer."	Mentzer et al. (2001, S. 4)
SCM	„Supply Chain Management beinhaltet die Planung, Steuerung und Kontrolle des gesamten Material- und Dienstleistungsflusses, einschließlich der damit verbundenen Informations- und Geldflüsse, innerhalb eines Netzwerkes von Unternehmungen, die im Rahmen von aufeinanderfolgenden Stufen der Wertschöpfungskette an der Entwicklung, Erstellung und Verwertung von Sachgütern und / oder Dienstleistungen partnerschaftlich zusammenarbeiten mit dem Ziel der Ergebnis- und Liquiditätsoptimierung - unter Beachtung von sozioökologischen Zielen."	Hahn (1999, S. 851)

Quelle: eigene Darstellung, nach [Mentzer et al., 2001, S. 4; Hahn, 1999, S. 851]

2.2.2. Aufgaben, Ziele und Konzepte des Supply Chain Managements

Obwohl in der Literatur eine Vielfalt an Definitionen zum SCM existiert, lassen sich dennoch einige gemeinsame Kernelemente beobachten, die in vielen Definitionen immer wieder vorkommen (Corsten & Gössinger, 2008, S. 109). Erstens wird vor allem die kooperative Zusammenarbeit der Teilnehmer eines Supply Chain Netzwerkes hervorgehoben. Zweitens ist der Ausgangspunkt der Steuerung der Supply Chain der Bedarf des Endkunden auf Basis von Daten der Verkaufsstellen. Drittens ist das SCM geschäftsprozessorientiert und strebt eine optimale und unternehmensübergreifende Gestaltung der Gesamtprozesse an (Cooper et al., 1997, S. 4; Seuring & Müller, 2009, S. 167; Corsten & Gössinger, 2008, S. 109).

Ein besonders wichtiger Aspekt für die Koordination im Supply Chain Netzwerk ist der Informationsfluss zwischen den Netzwerkteilnehmern, denn dieser hat direkten Einfluss auf die Produktionsplanung, die Bestandskontrolle im Lager sowie die Transportplanung aller Netzwerkteilnehmer (Lee, Padmanabhan & Whang, 1997, S. 546). Die Koordination kann nur verbessert werden, wenn alle Stufen des Netzwerkes Maßnahmen zur Erreichung gemeinsamer Ziele ergreifen (Chopra & Meindl, 2013, S. 262). Voraussetzung für den Aufbau eines SCM ist deshalb eine informationstechnische Verknüpfung aller Netzwerkteilnehmer, damit ein interorganisationaler Informationsfluss gesichert werden kann (Corsten & Gössinger, 2008, S. 110). Dabei können sogenannte Business Information Warehouse Systeme helfen, indem es alle relevanten Informationen zusammenfasst und den entsprechenden Netzwerkteilnehmern zur Verfügung stellt (Pokorný & Sokolowsky, 1999, S. 667).

Kommt es innerhalb der Supply Chain zu einem mangelnden Informationsfluss, so kann dies zum sogenannten Bullwhip-Effekt bzw. Peitschenschlag-Effekt führen (Seuring & Müller, 2009, S. 167). Dieser stellt das Problem einer Nachfrageverzerrung und -aufschaukelung im Supply Chain Netzwerk dar (Göpfert, 2004, S. 33). Kleine Schwankungen des Kundenbedarfs führen bei mangelnder Koordination, d. h. lokal begrenzten Informationen und lokalen Entscheidungen zwischen den Netzwerkteilnehmern auf jeder vorgelagerten Stufe des Supply Chain Netzwerkes zu einer immer größer werdenden Varianz der Bedarfsmengen. Ist jedem Netzwerkteilnehmer nur die Bedarfsmenge des direkten Nachfolgers bekannt, wird mit zunehmenden Abstand zum Endkunden die Gefahr der Fehleinschätzung der Nachfrage immer größer (Seuring & Müller, 2009, S. 167; Göpfert, 2004, S. 33–34). Die Lösung des Bullwhip-Effekts, also die „Synchronisation von Nachfrage und Angebot in interorganisationalen Wertschöpfungsketten", wird in der Literatur überwiegend als Hauptmotiv des SCM betont (Göpfert, 2004, S. 33). Nach Göpfert (2004, S. 35) sind spezifische Zielsetzungen des SCM vom Bullwhip-Effekt abgeleitet und sollen zu dessen Lösung beitragen. Dazu zählen eine konsequente Orientierung an der Kundennachfrage, die Einbindung von Kunden und Lieferanten, eine interorganisationale Sicht auf Ressourcen und Lagerbestände, Bestandsabbau, eine verbesserte Ausnutzung von Kapazitäten, Ressourcenflexibilisierung, Verbesserung der Lieferbereitschaft sowie wachsende Umsätze und Renditen (Göpfert, 2004, S. 35). Weiterhin nennen Busch und Dangelmaier (2004, S. 8) den Abbau von Informationsasymmetrien durch mehr Transparenz sowie höhere Kontinuität in den interorganisationalen Informations-, Material- und Geldflüssen. Im Hinblick auf die genannten Zielsetzungen kann konstatiert werden, dass das Hauptziel des SCM die Effizienzverbesserung der gesamten interorganisationalen Wertschöpfungskette ist.

Zur Erreichung der Ziele obliegen dem SCM strategische, taktische und operative Aufgaben (Corsten & Gössinger, 2008, S. 111). Auf der strategischen Ebene werden Entscheidungen getroffen, die einen langfristigen Effekt auf das Unternehmen haben (Simchi-Levi, Simchi-Levi & Kaminsky, 2004, S. 13). Im Rahmen des Supply Chain Designs werden auf dieser Ebene Planungs- und Koordinationsentscheidun-

gen bezüglich der Netzwerkstruktur getroffen, wie z. B. bezüglich der Anzahl, des Standorts oder die Kapazitäten von Warenlagern und Produktionsstandorten oder die Auswahl und Anzahl der Lieferanten (Busch & Dangelmaier, 2004, S. 7; Simchi-Levi et al., 2004, S. 13; Corsten & Gössinger, 2008, S. 111). Auf der taktischen, mittelfristigen Ebene steht die Nachfrage- und Lieferkettenplanung im Mittelpunkt (Corsten & Gössinger, 2008, S. 114). Hier werden Entscheidungen zur Lagerhaltung, Produktionsplanung oder Transportstrategien getroffen. Der operativen, kurzfristigen Ebene obliegen Entscheidungen des täglichen Geschäfts (Simchi-Levi et al., 2004, S. 13). Hier findet „die inhaltliche, mengenmäßige und zeitliche Abstimmung der Beschaffungs-, Produktions-, und Distributionsmengen bezogen auf den einzelnen Akteuren der Lieferkette“ statt (Corsten & Gössinger, 2008, S. 114). Es werden z. B. Transportwege, Durchlaufzeiten oder die LKW-Beladung auf Grundlage der taktischen Lieferkettenplanung, der Kundenaufträge und verfügbaren Ressourcen geplant (Simchi-Levi et al., 2004, S. 13; Corsten & Gössinger, 2008, S. 114).

Um diese SCM-Aufgaben zu erfüllen, werden in der Praxis verschiedene Konzepte eingesetzt (Busch & Dangelmaier, 2004, S. 7).[3] Das strategische Konzept Efficient Consumer Response (ECR) setzt direkt am Kunden an, um die Nachfrage effizient decken zu können (Corsten & Gössinger, 2008, S. 123). Das Ziel ist die Ausrichtung aller Supply Chain Netzwerkteilnehmer auf den maximalen Nutzen der Kunden bei gleichzeitiger Kostensenkung durch die Minimierung von Lieferzeiten im Absatzkanal, die Reduzierung von Lagerbeständen, die Vermeidung von Dopplungseffekten bei den Kosten für die Logistik sowie die Verbesserung des Logistikservices (Fischer & Städler, 1999, S. 349[4], zitiert nach Corsten & Gössinger, 2008, S. 124). Ein Teilkonzept des ECR ist das vom Produzenten betriebene Bestandsmanagement, das sogenannte Vendor Managed Inventory (VMI) (Baumgarten & Darkow, 2004, S. 101). Bei diesem Konzept ist der Lieferant für die Disposition des Eingangslagers des Abnehmers verantwortlich. Der wesentliche Vorteil für den Lieferanten liegt darin, dass dieser seine Produktion schnell anpassen und somit wirtschaftlichere Lose erreichen kann. Für den Abnehmer hat das VMI den Vorteil, dass neben dem Bestellaufwand auch die Lagerbestände reduziert werden, was wiederum zu niedrigeren Kapitalbindungskosten führt (Vahrenkamp, Kotzab & Siepermann, 2012, S. 219–220; Baumgarten & Darkow, 2004, S. 101–102).

Während das VMI-Konzept eher für Waren mit schwankendem Bedarf geeignet ist, ist das sogenannte Just-In-Time (JIT) Konzept bei Waren mit stetigem Bedarf vorteilhafter (Vahrenkamp et al., 2012, S. 220). JIT findet seinen Ursprung im sogenannten Toyota Production System und soll eine schlanke Produktion bzw. Lean Production ermöglichen. Das Ziel des JIT-Konzepts ist es, Produkte und Dienstleistungen zur richtigen Zeit, in den richtigen Mengen, zur richtigen Qualität und am richtigen Ort bereitzustellen. Dies kann nur ermöglicht werden, wenn die Produkte und Dienstleistungen nur dann hergestellt und geliefert werden, wenn ein Pull-Signal zum Kauf durch einen Kunden stattfindet (Cudney & Elrod, 2011, S. 6). Diese bedarfssynchrone Auslieferung von Waren ermöglicht demnach eine höhere Flexibilität bei der Produktionssteuerung in der Massenproduktion (Vahrenkamp et al., 2012, S. 219). Corbett und Klassen (2006, S. 12) argumentieren, dass SCM-Konzepte wie JIT und Lean Production bereits Konzepte eines SSCM sind. Durch die Reduzierung von Verschwendung und die einhergehenden Prozessverbesserungen können diese Konzepte bereits einen positiven Einfluss auf die ökologische Unternehmensperformance haben (Corbett & Klassen, 2006, S. 12). In den folgenden Abschnitten soll deshalb nach einer vorherigen Definition des SSCM und dessen Treibern näher auf die verschiedenen Praktiken des SSCM eingegangen werden.

2.3. Sustainable Supply Chain Management

Nachdem zuvor die Konzepte der Nachhaltigkeit, das unternehmerische Nachhaltigkeitsmanagement und das SCM vorgestellt wurden, werden in diesem Kapitel diese Konzepte zum SSCM zusammengeführt. Wie im vorherigen Abschnitt schon erwähnt, sind die Produktionsprozesse von Unternehmen rund um den Globus verteilt und die fokalen Unternehmen, Lieferanten und Kunden durch Informations-, Material- und Kapitalflüsse miteinander verknüpft. Dies mag zwar ökonomische Vorteile haben, aber mit jeder zusätzlichen Stufe im Produktionsprozess eines Produktes wächst die Gefahr einer höheren Belastung für Umwelt und Gesellschaft. Demzufolge sollten fokale Unternehmen des Supply Chain Netzwerkes auch für die ökologischen und sozialen Belastungen ihrer Lieferanten verantwortlich gemacht werden (Seuring & Müller, 2008, S. 1699). Im Folgenden wird eine für diese Arbeit relevante Definition des SSCM festgelegt. Anschließend werden Treiber des SSCM anhand von Organisationtheorien identifiziert.

2.3.1. Definition des Sustainable Supply Chain Managements

Obwohl das SCM in den letzten drei Dekaden umfassend in Theorie und Praxis untersucht wurde, haben die Diskussionen über Nachhaltigkeit im SCM erst Anfang der 2000er Jahre begonnen (Mitra & Datta, 2014, S. 2085). In Anbetracht des zunehmenden Drucks durch die Nachhaltigkeitsdebatte wird von Unternehmen erwartet, Konzepte zur Reduzierung von Auswirkungen ihrer Produkte und Dienstleistungen auf die Umwelt umzusetzen (Sarkis, 2001, S. 666). Unternehmen wird deshalb die Implementierung von SSCM empfohlen, um Anforderungen der Stakeholder zu erfüllen, die Wettbewerbsfähigkeit und Rentabilität zu erhöhen und gleichzeitig die ökologische und soziale Unternehmensperformance zu verbessern. Jedoch existiert ähnlich wie beim

[3] Zum Verständnis wird im Folgenden nur kurz auf die bedeutendsten Konzepte eingegangen, da eine ausführliche Erläuterung im Rahmen dieser Arbeit nicht zielführend hinsichtlich der Beantwortung der Forschungsfragen wäre.

[4] Fischer und Städler (1999): Efficient Consumer Response und zwischenbetriebliche Integration, in: Hippner, H.; Meyer, M.; Wilde, K.D. (Hrsg.): Computer Based Marketing: Das Handbuch zur Marketinginformatik, 2. Aufl., Vieweg+Teubner Verlag Wiesbaden 1999, S. 349–356.

SCM auch beim SSCM keine einheitliche Definition der Begrifflichkeit (Gualandris & Kalchschmidt, 2014, S. 93).

Das Konzept des SSCM bezieht den Nachhaltigkeitsaspekt in die Definition des SCM mit ein (Saeed, Waseek & Kersten, 2017, S. 162). Nach Schaltegger und Harms (2010, S. 6) stellt das SSCM die „inhaltliche Verknüpfung von Supply Chain Management und Nachhaltigkeitsmanagement“ dar. Es werden also bei der Optimierung der Supply Chain neben ökonomischen Gesichtspunkten auch die Auswirkungen auf Umwelt und Soziales berücksichtigt. Die bedeutendsten Definitionen des SSCM bieten Carter und Rogers (2008, S. 368) und Seuring und Müller (2008, S. 1700). Carter und Rogers (2008, S. 368) definieren SSCM als „the strategic, transparent integration and achievement of an organization's social, environmental, and economic goals in the systemic coordination of key interorganizational business processes for improving the long-term economic performance of the individual company and its supply chains.” Ähnlich ist die Definition von Seuring und Müller (2008, S. 1700). Die Autor*innen bezeichnen SSCM als „the management of material, information and capital flows as well as cooperation among companies along the supply chain while taking goals from all three dimensions of sustainable development, i.e., economic, environmental and social, into account which are derived from customer and stakeholder requirements” (Seuring & Müller, 2008, S. 1700). Bei der Definition von Carter und Rogers (2008, S. 368) dient die Integration dieser Ziele in die interorganisationalen Geschäftsprozesse vorwiegend einer langfristigen Verbesserung der ökonomischen Unternehmensperformance. Die langfristige Verbesserung der ökologischen und sozialen Unternehmensperformance wird hier jedoch nicht erwähnt (Pagell & Wu, 2009, S. 38). Eine umfassende Nachhaltigkeit der Supply Chain kann jedoch nur erreicht werden, wenn alle Teilnehmer zusammenarbeiten, um zum einen die drei Nachhaltigkeitsziele zu erreichen und zum anderen auch den Ansprüchen der Kunden und anderer Stakeholder gerecht zu werden (Saeed & Kersten, 2019, S. 3), was vor allem in der Definition von Seuring und Müller (2008, S. 1700) zur Geltung kommt. Weiterhin gehen Seuring und Müller (2008, S. 1700) nicht explizit auf die Verbesserung der ökonomischen Unternehmensperformance als oberstes Ziel des SSCM ein, sodass gemäß ihrer Definition angenommen werden kann, dass alle drei Dimensionen als gleichberechtigt angesehen werden. Da in der vorliegenden Arbeit der Zusammenhang von SSCM und allen drei Dimensionen der Unternehmensperformance untersucht werden soll, wird deshalb die Definition von Seuring und Müller (2008, S. 1700) herangezogen. Weiterhin geht diese Definition in ihren Teilaspekten mit den zuvor festgelegten Definitionen der Supply Chain (Mentzer et al., 2001, S. 4) und des SCM (Hahn, 1999, S. 851) einher.

2.3.2. Treiber des Sustainable Supply Chain Managements

Unternehmen und ihre Supply Chains stehen zunehmend unter Druck, denn sie werden nicht nur von internen, sondern auch von externen Stakeholdern getrieben, nachhaltige Praktiken in ihre Supply Chains zu integrieren (Hsu et al., 2013, S. 657; Delmas & Toffel, 2004, S. 219). Dieser interne und externe Druck auf Unternehmen wird in der Literatur häufig als SSCM Treiber definiert (Hsu et al., 2013, S. 663). Nach Saeed et al. (2017, S. 163) stellen diese Treiber sogenannte Motivatoren dar, die Unternehmen dazu anregen sollen, nachhaltige Praktiken in das SCM zu integrieren. In der Literatur werden verschiedene Organisationstheorien betrachtet, um zu verstehen, weshalb Unternehmen nachhaltige Praktiken in ihr SCM einbauen (Varsei, Soosay, Fahimnia & Sarkis, 2014, S. 244). Allgemein helfen Organisationstheorien dabei, das Verhalten und die Strukturen von Unternehmen zu verstehen (Sarkis, Zhu & Lai, 2011, S. 2). Varsei et al. (2014, S. 245) identifizierten vier Theorien, um herauszufinden, welche Faktoren die Integration nachhaltiger Praktiken in einem Supply Chain Netzwerk antreiben und ermöglichen: Resourced-Based-Theory (RBT), Institutional Theory, Stakeholder Theory sowie die Social Network Theory (SNT).

Nach der RBT sind wertvolle, seltene sowie nicht imitierbare und nicht substituierbare Ressourcen und Fähigkeiten die Basis für Wettbewerbsvorteile von Unternehmen. Ressourcen und Fähigkeiten sind z. B. alle Vermögenswerte, Prozesse, Kompetenzen und Kenntnisse, die ein Unternehmen kontrolliert, um Strategien zu entwickeln und gleichzeitig die Wettbewerbsfähigkeit zu erhöhen (Barney, 1991, S. 99–101). Hart (1995, S. 991) erweitert die RBT um den Nachhaltigkeitsaspekt, da es wahrscheinlich sei, dass zukünftig die Wettbewerbsvorteile auch auf umweltverträglichen Unternehmensressourcen beruhen. Dies bezeichnet der Autor als „Natural-Resource-Based view of the firm” (NRBV) (Hart, 1995, S. 991). In der Literatur gibt es seit geraumer Zeit Anstrengungen den Zusammenhang von RBT und SCM zu untersuchen (Gold, Seuring & Beske, 2010; Rao & Holt, 2005; Sarkis et al., 2011). Da sich das äußere Umfeld von Unternehmen stetig ändert, müssen diese ihre Ressourcen und Fähigkeiten ständig anpassen und neu ausrichten, um weiterhin von Wettbewerbsvorteilen profitieren zu können (Gold et al., 2010, 232). Sogenannte Dynamic Capabilities erleichtern diese stetige Anpassung an äußere Einflüsse. Nach C. L. Wang und Ahmed (2007, S. 35) dienen Dynamic Capabilities dem Unternehmen, seine Ressourcen gemäß den äußeren Umständen stetig neu anzupassen, indem es z. B. seine Kernkompetenzen neu ausrichtet, um weiterhin Wettbewerbsvorteile zu generieren. Demnach können Unternehmen ihr Supply Chain Netzwerk nutzen, um auf Ressourcen und Fähigkeiten der Netzwerkpartner zuzugreifen (Gold et al., 2010, 232). Eine effektive Nutzung gemeinsamer Ressourcen und Fähigkeiten kann somit zu Wettbewerbsvorteilen führen (Varsei et al., 2014, S. 246). In Bezug auf SSCM stellen diese interorganisationalen Ressourcen und Fähigkeiten nachhaltige Praktiken dar, die es der Supply Chain möglich machen, ökologisch und sozial nachhaltig zu werden (Sarkis et al., 2011, S. 8). Diese nachhaltigen Praktiken verbessern nicht nur das Image des fokalen Unternehmens, sondern auch das der Produkte und Dienstleistungen (Shang, Lu & Li, 2010, S. 1224).

Als zweite Organisationstheorie nennen Varsei et al. (2014, S. 245) die Institutional Theory. Diese beschreibt, wie Institutionen (z. B. Regierungen oder Medien) Druck auf

Unternehmen ausüben, um deren organisatorische Verhaltensweisen zu beeinflussen (Oliver, 1991, S. 145; Varsei et al., 2014, S. 245). In Bezug auf SSCM wird die Institutional Theory verwendet, um zu erklären, weshalb SSCM Praktiken von Unternehmen implementiert werden (Saeed & Kersten, 2019, S. 3). Die nachhaltige Ausrichtung des SCM wird in der Literatur vom normativen, zwanghaften (engl. coercive) und mimetischen externen Druck beeinflusst (DiMaggio & Powell, 1983, S. 150; Hsu et al., 2013, S. 661). Normativer Druck entsteht durch soziale Verpflichtungen und wird durch externe Stakeholder, die ein bestimmtes Interesse am Unternehmen haben, wie z. B. Gewerkschaften, NGOs, Kunden oder Lieferanten, ausgeübt (Saeed & Kersten, 2019, S. 3–4; Zhu & Sarkis, 2007, S. 4335). Nach Ball und Craig (2010, S. 291–292) führt normativer Druck dazu, dass Unternehmen umweltbewusster und sozialer handeln. Zwanghafter Druck wird als einflussreichster Druck auf Unternehmen angesehen. Beispiele hierfür sind Regierungen oder Aufsichtsbehörden, die durch Regularien wie Umweltvorschriften Unternehmen dazu drängen, SSCM Praktiken umzusetzen, um den negativen Einfluss auf die Umwelt und Gesellschaft zu verringern (Rivera, 2004, S. 782; Zhu & Sarkis, 2007, S. 4335). Mimetischer Druck tritt auf, sobald Wettbewerber eines Unternehmens erfolgreicher sind (DiMaggio & Powell, 1983, S. 151). Wenn Wettbewerber eines Unternehmens also bereits SSCM Praktiken erfolgreich anwenden, wird das fokale Unternehmen dazu getrieben, auch solche Praktiken umzusetzen (Saeed & Kersten, 2019, S. 4). Dieser institutionelle Druck veranlasst Unternehmen und deren Supply Chains dazu, sozial- und umweltbewusste Praktiken anzuwenden (Varsei et al., 2014, S. 246–247).

Als dritte Organisationstheorie wird die Stakeholder Theory genannt, welche direkt an die Institutional Theory anknüpft (Varsei et al., 2014, S. 245). Freeman (2010, S. 46) definiert Stakeholder als „any group or individual who can affect or is affected by the achievement of the organization's objectives." Dies lässt sich auch auf Nachhaltigkeitsinitiativen von Unternehmen übertragen (Varsei et al., 2014, S. 246). Svensson, Ferro, Hogevold, Padin und Sosa Varela (2018, S. 20) identifizierten fünf Gruppen von Stakeholdern in Bezug auf die Umsetzung von SSCM Praktiken des fokalen Unternehmens in einem Supply Chain Netzwerk: Upstream Stakeholder (z. B. Lieferanten), das fokale Unternehmen (z. B. Management, Aufsichtsrat, Mitarbeiter), Downstream Stakeholder (z. B. Groß- und Einzelhändler), Markt Stakeholder (z. B. Endkunden) sowie gesellschaftliche Stakeholder (z. B. Regierung, NGO, Gesellschaft). Unternehmen sind gezwungen die Interessen dieser Stakeholder zu erfüllen, damit die Rentabilität der Geschäftstätigkeit sichergestellt wird (Varsei et al., 2014, S. 247). Da sich die Stakeholder neben der Erreichung ökonomischer Ziele aber auch zunehmend um Umwelt- und Sozialaspekte sorgen, ist die Implementierung von SSCM Praktiken nach dieser Interpretation der Stakeholder Theory ratsam (Golicic & Smith, 2013, S. 92).

Als letzte Organisationstheorie nennen Varsei et al. (2014, S. 245) die SNT. Nach Chabowski, Mena und Gonzalez-Padron (2011, S. 57–58) ist ein soziales Netzwerk im Business-Kontext eine Gruppe von Unternehmen, die durch eine Reihe von Abhängigkeiten und Beziehungen miteinander verbunden sind, wobei diese Abhängigkeiten und Beziehungen durch Knoten und Linien grafisch dargestellt werden können. Gemäß der SNT kann das in Kapitel 2.2.1 dargestellte Supply Chain Netzwerk somit auch als soziales Netzwerk von Unternehmen verstanden werden. Das Supply Chain Netzwerk besteht aus miteinander verbundenen Unternehmen, deren Erfolg von interorganisational integrierten Geschäftsprozessen und der kollaborativen Performance der einzelnen Netzwerkteilnehmer abhängt. Zudem kann ein effektiver Informationsfluss im Supply Chain Netzwerk die Umsetzung von SSCM Praktiken fördern (Varsei et al., 2014, S. 247). Auf Basis der vier vorgestellten Organisationstheorien kann festgestellt werden, dass die Motivation zur Umsetzung von SSCM Praktiken sowohl von internen als auch von externen Quellen getrieben wird (Delmas & Toffel, 2004, S. 210). Die Tabelle 2 stellt die Treiber noch einmal übersichtlich dar.

2.4. Verknüpfung von Sustainable Supply Chain Management mit der Unternehmensperformance

In der Literatur wird seit geraumer Zeit vor allem der Zusammenhang von GSCM Praktiken auf die ökologische und ökonomische Unternehmensperformance diskutiert (Zhu & Sarkis, 2004; Zhu, Sarkis & Geng, 2005). Aber auch soziale Praktiken stehen immer öfter im Zentrum wissenschaftlicher Untersuchungen (Carter & Jennings, 2002; Das, 2018; J. Wang & Dai, 2018), wobei der Einfluss auf die soziale Unternehmensperformance aber häufig noch vernachlässigt wird (Panigrahi et al., 2019, S. 1027). Da das Ziel der vorliegenden Arbeit die Analyse des Effekts von SSCM Praktiken auf die gesamte TBL von Unternehmen ist, wird in diesem Abschnitt der Zusammenhang von SSCM Praktiken und der Unternehmensperformance hergestellt. Dazu werden in der Literatur relevante SSCM Praktiken auf Grundlage der in Kapitel 2.3.2 vorgestellten Theorien erläutert. Da sich besonders die Ressourcentheorien RBT und NRBV mit den Beziehungen zwischen den Fähigkeiten und Wettbewerbsvorteilen von Unternehmen befassen, bilden diese Theorien die geeignete Grundlage zur Herleitung von Zusammenhängen zwischen SSCM Praktiken und der Unternehmensperformance (Golicic & Smith, 2013, S. 82). Weiterhin werden diese Zusammenhänge anhand der Ergebnisse früherer Studien untersucht und in Hypothesen dargestellt. Anschließend werden die Hypothesen zu einem theoretischen Forschungsmodell zusammengefügt, das als Grundlage für die durchzuführende Befragung und die anschließende Datenanalyse dienen wird. Anhand einschlägiger Literatur konnten jeweils vier etablierte interne und externe SSCM Praktiken identifiziert werden (J. Wang & Dai, 2018, S. 4), die im Folgenden nacheinander näher erläutert werden.

2.4.1. Herleitung der Hypothesen

Internes Umweltmanagement

Tabelle 2: SSCM Treiber

Theorie	SSCM Treiber
Resourced-Based-Theory	• Wettbewerbsvorteile durch den Austausch nachhaltiger Ressourcen und Fähigkeiten zwischen Supply Chain Partnern
Institutional Theory	• Externe Stakeholder, die ein gewisses Interesse am Unternehmen haben (normativ) • Staatliche Institutionen, Regularien und Gesetze (zwanghaft) • Wettbewerber (mimetisch)
Stakeholder Theory	• Erfüllung der Interessen aller Stakeholder
Social Network Theory	• Erfolg des Supply Chain Netzwerkes (Social Network)

Quelle: eigene Darstellung, nach (Varsei et al., 2014, S. 246–247)

Als erste Praktik ist das interne Umweltmanagement (IU) zu nennen (Zhu & Sarkis, 2004, S. 267). IU bezieht sich u. a. auf unternehmensinterne Werte und ethische Standards hinsichtlich des Engagements zum Schutz der Umwelt (Chan, He, Chan & Wang, 2012, S. 623). In der Literatur herrscht Einigkeit darüber, dass IU Praktiken die Unternehmensperformance verbessern können (Yildiz Çankaya & Sezen, 2019, S. 114; J. Wang & Dai, 2018, S. 10). Die Einführung von Umweltmanagementsystemen (UMS) wie die ISO-Normreihe 14000 führt nach Klassen und Whybark (1999, S. 611) z. B. zu weniger Ausstoß von umweltgefährdenden Schadstoffen. Es liegt also nahe, dass IU Praktiken einen positiven Einfluss auf die ökologische Unternehmensperformance haben. J. Wang und Dai (2018, S. 10) haben dies in ihrer empirischen Studie bei 172 chinesischen Firmen belegen können. Auch Zhu und Sarkis (2004, S. 282) können diesen positiven Effekt anhand einer weiteren chinesischen Studie bestätigen. Das (2018, S. 5788) kommt in seiner Studie in indischen Unternehmen auf dasselbe Ergebnis. Demnach kann folgende Hypothese aufgestellt werden:

> **Hypothese 1a:** *IU hat einen positiven Einfluss auf die ökologische Unternehmensperformance.*

Die Integration von IU in Unternehmen kann dazu führen, dass der Verbrauch von Ressourcen reduziert, die Stakeholder Beziehungen verbessert und das Image des Unternehmens gesteigert wird. Der Einsatz umweltfreundlicher Materialien und Verfahren kann weiterhin zu effizienteren Produktionsprozessen führen. Dadurch kann wiederum der Umsatz gesteigert und die Produktionskosten können gesenkt werden (J. Wang & Dai, 2018, S. 5). Sowohl Zhu und Sarkis (2004, S. 277), Zhu et al. (2005, S. 460) als auch Rao und Holt (2005, S. 912) konnten einen positiven Zusammenhang zwischen IU und der ökonomischen Unternehmensperformance erkennen. Auf Basis dieser Ergebnisse lässt sich die nachfolgende Hypothese aufstellen:

> **Hypothese 1b:** *IU hat einen positiven Einfluss auf die ökonomische Unternehmensperformance.*

Wird IU in einem Unternehmen umgesetzt, so wird z. B. durch saubere Produktionsprozesse die Emission von Schadstoffen gesenkt. Dies kann zum einen die Arbeitsbedingungen der Mitarbeiter*innen verbessern, zum anderen hat dies einen positiven Einfluss auf das direkte Umfeld des fokalen Unternehmens. Darüber hinaus kann das soziale Ansehen erhöht werden (J. Wang & Dai, 2018, S. 5). Die Autor*innen fanden in ihrer Studie bei chinesischen Unternehmen heraus, dass die Integration von IU einen positiven Effekt auf die soziale Unternehmensperformance hat (J. Wang & Dai, 2018, S. 10). Auch Yildiz Çankaya und Sezen (2019, S. 113) können diesen positiven Zusammenhang anhand ihrer Studie in türkischen Unternehmen bestätigen. Deshalb kann folgende Hypothese aufgestellt werden:

> **Hypothese 1c:** *IU hat einen positiven Einfluss auf die soziale Unternehmensperformance.*

Nachhaltiges Produkt- und Prozessdesign

Aspekte der Nachhaltigkeit sollten sowohl im Produkt- als auch im Prozessdesign integriert werden (Thun & Müller, 2009, 120). Als zweite SSCM Praktik ist deshalb nachhaltiges Produkt- und Prozessdesign (NPP) zu nennen. Es gibt eine Reihe von Maßnahmen, um Produktionsprozesse und Produkte nachhaltiger zu gestalten. Dazu gehören z. B. der Einsatz umweltfreundlicher Rohstoffe, die Durchführung von Lebenszyklusanalysen zur ökologischen Evaluierung der Produkte, die Optimierung des Produktionsprozesses zur Reduzierung von Abfällen und Emissionen oder der Einsatz neuer Technologien zur Einsparung von Energie und Wasser, was wiederum zu Kosteneinsparungen führen kann (Rao & Holt, 2005, S. 902–903; Hsu et al., 2013, S. 673). Zudem können nach der RBT und NRBV mit nachhaltigen Innovationen im Produktionsprozess Vorteile gegenüber Wettbewerbern generiert werden (Kleindorfer, Singhal & Wassenhove, 2005, S. 485). Paulraj, Chen und Blome (2017, S. 253) kamen in ihrer Studie über 259 Unternehmen in Deutschland zu der Erkenntnis, dass NPP einen positiven Effekt sowohl auf die ökologische als auch auf die ökonomische Unternehmensperformance hat. Auf Basis dieser Erkenntnisse werden folgende Hypothesen aufgestellt:

Hypothese 2a: *NPP hat einen positiven Einfluss auf die ökologische Unternehmensperformance.*

Hypothese 2b: *NPP hat einen positiven Einfluss auf die ökonomische Unternehmensperformance.*

Ob NPP auch einen Einfluss auf die soziale Unternehmensperformance hat, wird in der Literatur nicht explizit herausgearbeitet. J. Wang und Dai (2018, S. 19) haben die zu NPP äquivalente Variable „Eco-Design" unter dem Konstrukt „Internal Management" zusammengefasst, welches aber insgesamt einen positiven Effekt auf die soziale Unternehmensperformance hat, weshalb angenommen werden kann, dass auch NPP diesen Effekt auslöst. Zudem lässt sich auch hier argumentieren, dass sauberere Produktionsprozesse und Produkte das Image des Unternehmens sowie das Wohlbefinden der Mitarbeiter und der umliegenden Gemeinden grundsätzlich verbessern können (J. Wang & Dai, 2018, S. 5; Yildiz Çankaya & Sezen, 2019, S. 113). Aus diesen Gründen lässt sich folgende Hypothese aufstellen:

Hypothese 2c: *NPP hat einen positiven Einfluss auf die soziale Unternehmensperformance.*

Investment Recovery

Investment Recovery (IR) ist schon seit Jahren in traditionelle Geschäftsprozesse integriert und ist auch im Bereich SSCM eine etablierte Praktik (Yildiz Çankaya & Sezen, 2019, S. 101), bei der überschüssige, defekte oder gebrauchte Lagerbestände an Produkten, Materialien oder Schrott wiederverwendet, recycelt oder verkauft werden (Zhu & Sarkis, 2004, S. 268). Mithilfe von IR soll der höchstmögliche Wert aus diesen Gegenständen erzielt werden, um einerseits Verschwendung zu vermeiden und andererseits finanzielle Mittel einzusparen (Ayres, Ferrer & van Leynseele, 1997, S. 558). Auf Basis dieser theoretisch hergeleiteten Kenntnisse kann demnach angenommen werden, dass IR einen positiven Effekt auf die ökologische Unternehmensperformance haben kann. Zhu und Sarkis (2004, S. 281) kamen in ihrer empirischen Studie in chinesischen Unternehmen zu dem Ergebnis, dass IR einen hoch signifikanten Einfluss auf die ökologische Unternehmensperformance hat. Auch Esfahbodi, Zhang, Watson und Zhang (2017, S. 24), Yildiz Çankaya und Sezen (2019, S. 111) und Y. Lu, Zhao, Xu und Shen (2018, S. 20) erkannten einen positiven Zusammenhang. Es wird somit folgende Hypothese aus diesen Ergebnissen abgeleitet:

Hypothese 3a: *IR hat einen positiven Einfluss auf die ökologische Unternehmensperformance.*

Die Eliminierung von Verschwendung führt zur Reduzierung von Kosten, sodass die ökonomische Unternehmensperformance verbessert werden könnte (Y. Lu et al., 2018, S. 6). Zhu und Sarkis (2004, S. 281) konnten einen schwach positiven Effekt von IR auf die ökonomische Unternehmensperformance bestätigen. Auch Y. Lu et al. (2018, S. 20) können in chinesischen Unternehmen einen positiven Einfluss empirisch belegen. Deshalb kann folgende Hypothese aufgestellt werden:

Hypothese 3b: *IR hat einen positiven Einfluss auf die ökonomische Unternehmensperformance.*

Wenn ein Unternehmen IR Praktiken umsetzt, kann dies zur Verbesserung des Unternehmensumfelds und somit zu einer höheren Mitarbeiterzufriedenheit und einer besseren Reputation des Unternehmens führen (Y. Lu et al., 2018, S. 21). Sowohl Yildiz Çankaya und Sezen (2019, S. 113) als auch Y. Lu et al. (2018, S. 21) haben einen positiven Effekt von IR auf die soziale Unternehmensperformance empirisch finden können. Somit lässt sich folgende Hypothese aufstellen:

Hypothese 3c: *IR hat einen positiven Einfluss auf die soziale Unternehmensperformance.*

*Soziale Verantwortung gegenüber Mitarbeiter*innen und Gesellschaft*

Ein Unternehmen kann es sich nicht erlauben, die Arbeitsbedingungen, Sicherheit, Gesundheit und die Bildung seiner Mitarbeiter*innen und die der umliegenden Gemeinden zu missachten (Das, 2017, S. 1348). Mit der SSCM Praktik „soziale Verantwortung gegenüber Mitarbeiter*innen und Gesellschaft" (SV) wird auf sozialspezifische Aspekte im SSCM eingegangen. SV Praktiken sind in der Literatur häufig unter CSR zu finden (Carter & Jennings, 2002, S. 145). Soziale Verantwortung gegenüber Mitarbeiter*innen können Unternehmen z. B. durch die Bereitstellung fairer Löhne und Gehälter, ein sicheres, gesundes und positives Arbeitsumfeld, Gesundheitsleistungen, ausreichend Urlaub, flexible Arbeitszeitmodelle und persönliche Entwicklungschancen, wie z. B. Weiterbildungsprogramme und Schulungen gewährleisten (Welford & Frost, 2006, S. 174; R. X. Lu, Lee & Cheng, 2012, S. 164; Zhu, Liu & Lai, 2016, S. 420). Soziale Verantwortung gegenüber der Gesellschaft bedeutet z. B. die Schaffung von Arbeitsplätzen, Bereitstellung und Unterstützung von Gesundheits- und Bildungseinrichtungen oder Investitionen in soziale Projekte (R. X. Lu et al., 2012, S. 164; Zhu et al., 2016, S. 420).

Welford und Frost (2006, S. 173) argumentieren, dass CSR Praktiken die Fehlzeiten verringern, die Arbeitsmoral erhöhen, die Loyalität der Beschäftigten zum Unternehmen verbessern und somit die Produktivität des Unternehmens steigern. Zhu et al. (2016, S. 423–424) zeigten in einer empirischen Studie in chinesischen Firmen, dass CSR Praktiken mit dem Schwerpunkt auf Corporate Governance und die Einhaltung von Menschenrechten einen signifikanten positiven Einfluss auf die soziale Unternehmensperformance haben. Weiterhin fanden die Autor*innen heraus, dass die ökonomische Unternehmensperformance vor allem durch gute Arbeitsbedingungen, gesellschaftliches und politisches Engagement sowie nachhaltiges Lieferantenmanagement positiv beeinflusst wird. Ein spezifischer Zusammenhang zwischen SV und der ökologischen Unternehmensperformance

konnte in der Literatur nicht identifiziert werden. Dennoch kann die Vermutung angestellt werden, dass auch SV einen positiven Zusammenhang aufweist, denn in Weiterbildungsmaßnahmen und Schulungen der Beschäftigten können auch umweltspezifische Aspekte behandelt werden (Sammalisto & Brorson, 2008, S. 299). Auf Basis dieser Erkenntnisse lassen sich drei weitere Hypothesen aufstellen:

> **Hypothese 4a:** *SV hat einen positiven Einfluss auf die ökologische Unternehmensperformance.*
>
> **Hypothese 4b:** *SV hat einen positiven Einfluss auf die ökonomische Unternehmensperformance.*
>
> **Hypothese 4c:** *SV hat einen positiven Einfluss auf die soziale Unternehmensperformance.*

Green Distribution

Green Distribution (GD) stellt eine externe SSCM Praktik dar. GD setzt sich aus nachhaltiger Verpackung und Logistik zusammen (Ninlawan, Seksan, Tossapol & Pilada, 2010, S. 2). Die meisten Produkte werden derzeit in Verpackungen geliefert, um Schäden am Produkt zu verhindern oder den Transport zu erleichtern (Rao & Holt, 2005, S. 904). Verpackungseigenschaften wie Material, Größe und Form wirken sich auf den Transport der darin enthaltenen Produkte aus und können zu hohen Mengen an Abfall führen (Ho, Shalishali, Tseng & Ang, 2009, S. 27; Rao & Holt, 2005, S. 904). Aus diesem Grund haben viele Länder inzwischen Rechtsvorschriften festgelegt, um den Verbrauch von Verpackungsmaterial zu senken (Rao & Holt, 2005, S. 904). Werden umweltfreundliche und recyclebare Materialien für Verpackungen verwendet und das Verpackungsdesign optimiert, führt dies zu weniger Abfällen sowie geringeren Entsorgungs-, Transport- und Lagerkosten (Wu & Dunn, 1995, S. 29). Weiterhin kann der Transport durch die Zusammenlegung von Aufträgen und die Routenoptimierung zu geringerem Energieverbrauch und weniger CO_2-Emissionen führen (Kafa, Hani & El Mhamedi, 2013, S. 72). Auch das öffentliche Interesse richtet sich immer mehr auf nachhaltige und sichere Verpackungen. Durch dessen Umsetzung können Unternehmen nicht nur die Bedürfnisse der Kunden nach nachhaltig und sicher verpackten Produkten befriedigen, sondern auch das Image der Produkte und des Unternehmens verbessern (Zailani, Jeyaraman, Vengadasan & Premkumar, 2012, S. 338).

Basierend auf diesen theoretischen Überlegungen kann davon ausgegangen werden, dass GD womöglich einen positiven Zusammenhang mit allen drei Dimensionen der Unternehmensperformance aufweist. Diese positiven Zusammenhänge können z. B. Zailani et al. (2012, S. 338) anhand einer Survey über 400 verarbeitende und herstellende Unternehmen in Malaysia bestätigen. Demzufolge lassen sich drei weitere Hypothesen aufstellen:

> **Hypothese 5a:** *GD hat einen positiven Einfluss auf die ökologische Unternehmensperformance.*
>
> **Hypothese 5b:** *GD hat einen positiven Einfluss auf die ökonomische Unternehmensperformance.*
>
> **Hypothese 5c:** *GD hat einen positiven Einfluss auf die soziale Unternehmensperformance.*

Lieferantenauswahl und -überwachung

Viele Unternehmen aus Industrienationen haben in den letzten Jahrzehnten die Herstellung von Produkten und Teilkomponenten sowie Dienstleistungen ausgelagert, um Kosten zu sparen und sich auf ihre Kernkompetenzen zu konzentrieren (Ehrgott, Reimann, Kaufmann & Carter, 2013, S. 131; Krause, Scannell & Calantone, 2000, S. 33). Infolgedessen sind heutzutage viele Zulieferer in Schwellenländern angesiedelt (Busse, Schleper, Niu & Wagner, 2016, S. 442). Nicht alle Lieferanten zeigen dabei ethisches Verhalten in Bezug auf umwelt- und sozialkritische Aspekte (J. Wang & Dai, 2018, S. 6). Besonders in Schwellenländern herrschen häufig schlechte Nachhaltigkeitsbedingungen bei den Lieferanten (Busse, 2016, S. 29). Um die Supply Chain hinsichtlich des Umweltschutzes und sozialer Verantwortung zu stärken, haben sich in vielen Unternehmen aufwendige Praktiken zur Lieferantenauswahl und -überwachung (LAÜ) etabliert (J. Wang & Dai, 2018, S. 6), da die reine Lieferantenauswahl hinsichtlich des besten Preises heutzutage nicht mehr akzeptabel ist (Sarkis & Talluri, 2002, S. 18). Es sollten neben dem Preis auch Kriterien, wie z. B. die finanzielle Situation des Lieferanten, die strategische Ausrichtung oder die Produktionskapazitäten und -qualitäten berücksichtigt werden (Ellram, 1990, S. 12). Im Zuge der Nachhaltigkeitsdebatte beziehen immer mehr Unternehmen auch umwelt- und sozialspezifische Aspekte in die LAÜ mit ein. Eine typische Praktik ist die Festlegung von Kriterien zur Bewertung der Lieferanten anhand ihrer ökologischen und sozialen Performance (Gualandris & Kalchschmidt, 2014, S. 93). Bewertungskriterien sind z. B., dass die Lieferanten Zertifizierungen für Umwelt- oder Sozialverantwortlichkeit, wie ISO 14000 oder SA 8000, vorweisen können (Gavronski, Klassen, Vachon & Nascimento, 2011, S. 882; J. Wang & Dai, 2018, S. 6). Weiterhin ist die Bewertung der Lieferanten Bestandteil eines dauerhaften Monitorings, mit dem Anreize aber auch Sanktionen verbunden sind. Fallen die Ergebnisse der Lieferantenevaluierung negativ aus (Delmas & Montiel, 2009, S. 195–196), so können diese Partnerschaften auch beendet werden. Dies kommt aufgrund zusätzlicher Transaktionskosten bei der Akquisition neuer Lieferanten aber eher selten vor (Hansen, Harms & Schaltegger, 2011, S. 92).

Bereits aus den theoretisch hergeleiteten Erkenntnissen kann vermutet werden, dass sich LAÜ nach nachhaltigen Bewertungskriterien positiv auf alle drei Dimensionen der Unternehmensperformance auswirkt. Gimenez, Sierra und Rodon (2012, S. 155) konnten in ihrer Studie über herstellende Unternehmen in 19 Industrienationen zwar einen minimal positiven Effekt auf die ökologische, ökonomische und soziale Unternehmensperformance erkennen, welcher aber nicht

als statistisch signifikant einzustufen ist. Zu einem ähnlichen Ergebnis kamen auch J. Wang und Dai (2018, S. 13). Um diesen Zusammenhang im dennoch zu überprüfen, können aus den theoretischen Überlegungen folgende Hypothesen aufgestellt werden:

Hypothese 6a: *LAÜ hat einen positiven Einfluss auf die ökologische Unternehmensperformance.*

Hypothese 6b: *LAÜ hat einen positiven Einfluss auf die ökonomische Unternehmensperformance.*

Hypothese 6c: *LAÜ hat einen positiven Einfluss auf die soziale Unternehmensperformance.*

Zusammenarbeit mit Lieferanten

Mit der Absicht den größten Nutzen aus ihren Nachhaltigkeitsanstrengungen zu ziehen, müssen Unternehmen die Mitglieder des Supply Chain Netzwerkes in ihre Prozesse integrieren (Walton, Handfield & Melnyk, 1998, S. 3). Nach Gualandris und Kalchschmidt (2014, S. 93) besteht die Zusammenarbeit mit Lieferanten (ZL) aus Maßnahmen, wie die gemeinsame Entwicklung eines nachhaltigen Produkt- und Prozessdesigns, die Reduzierung von Abfällen bei der Distribution von Materialien und Produkten, die Förderung des Informationsaustauschs zu Umwelt- und Sozialthemen zwischen Lieferanten und dem fokalen Unternehmen sowie das gemeinsame Management von Umwelt- und Sozialrisiken. Diese Zusammenarbeit erfordert aber auch, dass fokale Unternehmen ihren Lieferanten Ressourcen bereitstellen, um in kooperativen Projekten Umwelt- und Sozialprobleme zu bewältigen (Vachon & Klassen, 2006, S. 799). Zwar entstehen durch solche gemeinsamen Projekte Koordinationskosten; Informationskosten und Unsicherheiten können aber verringert werden (Hansen et al., 2011, S. 92).

Entsprechend des NRBV nach Hart (1995, S. 991) konnten Surroca, Tribó und Waddock (2010, S. 482) in einer Studie über 599 Unternehmen in großen Industrienationen einen positiven Zusammenhang zwischen immateriellen Unternehmenswerten, wie z. B. Lieferantenpartnerschaften, mit der ökonomischen Unternehmensperformance feststellen. Gimenez et al. (2012, S. 155) konnten neben dem positiven Effekt auf die ökologische Unternehmensperformance auch positive Effekte auf die ökologische und soziale Unternehmensperformance aufzeigen. J. Wang und Dai (2018, S. 13) erkannten einen positiven Zusammenhang von ZL auf die ökologische Unternehmensperformance. Somit können folgende drei Hypothesen aufgestellt werden:

Hypothese 7a: *ZL hat einen positiven Einfluss auf die ökologische Unternehmensperformance.*

Hypothese 7b: *ZL hat einen positiven Einfluss auf die ökonomische Unternehmensperformance.*

Hypothese 7c: *ZL hat einen positiven Einfluss auf die soziale Unternehmensperformance.*

Nachhaltige Beschaffung

Die Beschaffungsfunktion in Unternehmen hat im Laufe der Nachhaltigkeitsdebatte der letzten Jahrzehnte Strategien entwickeln müssen, um nachhaltige Aspekte bei der Beschaffung von Materialien und Produkten zu berücksichtigen (Min & Galle, 1997, S. 10). Dennoch dürfen auch traditionelle Beschaffungskriterien wie Kosten, Qualität und Lieferzeiten nicht außer Acht gelassen werden (Grün & Brunner, 2008, S. 68). Kriterien für nachhaltige Beschaffung (NB) sind z. B. die Berücksichtigung von Öko-Kennzeichnungen der zu beschaffenden Produkte und Materialien, die Wiederverwendbarkeit dieser, das Achtgeben auf umweltfreundliche Verpackungen sowie, dass die zu beschaffenden Produkte keine umweltschädlichen, giftigen oder gefährlichen Stoffe enthalten (Zhu, Sarkis & Lai, 2008, S. 271; Min & Galle, 1997, S. 11; Hsu et al., 2013, S. 673). Esfahbodi et al. (2017, S. 22) zeigten in ihrer empirischen Studie in Großbritannien, dass NB einen signifikanten und positiven Einfluss auf die ökologische und ökonomische Unternehmensperformance hat. Yildiz Çankaya und Sezen (2019, S. 111) sowie Zhu, Sarkis und Lai (2007, S. 1047) konnten jedoch keinen signifikanten Zusammenhang über alle drei Dimensionen erkennen. Aufgrund der theoretischen Überlegungen und der Ergebnisse von Esfahbodi et al. (2017, S. 22) werden trotz der widersprüchlichen Erkenntnisse in der Literatur folgende drei Hypothesen aufgestellt:

Hypothese 8a: *NB hat einen positiven Einfluss auf die ökologische Unternehmensperformance.*

Hypothese 8b: *NB hat einen positiven Einfluss auf die ökonomische Unternehmensperformance.*

Hypothese 8c: *NB hat einen positiven Einfluss auf die soziale Unternehmensperformance.*

Interaktionen zwischen den drei Dimensionen der Unternehmensperformance

SSCM Praktiken sind wichtige Ressourcen für Unternehmen, um sich durch den Aufbau von Wettbewerbsvorteilen von ihren Wettbewerbern abzuheben. Sie sind dadurch u. a. in der Lage, den Energie- und Materialverbrauch, Abfälle, Abwasser und Emissionen zu verringern, Umwelt- und Arbeitsunfälle zu vermeiden und somit ihre Reputation zu erhöhen. Im Vergleich zu ihren Wettbewerbern haben Unternehmen mit einer besseren Umwelt- und Sozialperformance ein höheres Ansehen in ihrer Branche und genießen daher höhere Kundenzufriedenheit und Loyalität ihrer Stakeholder (J. Wang & Dai, 2018, S. 7). Somit kann vermutet werden, dass die ökologische und soziale Unternehmensperformance einen positiven Effekt auf die ökonomische Unternehmensperformance haben. Gestützt wird diese Vermutung durch die Ergebnisse von Green Jr., Zelbst, Meacham und Bhadauria (2012, S. 299). Die Autor*innen konnten in ihrer Befragung bei 159 Operations Managern in US-Unternehmen

einen signifikanten und positiven Effekt der ökologischen auf die ökonomische Unternehmensperformance erkennen. J. Wang und Dai (2018, S. 13) haben signifikant positive Effekte der ökologischen und sozialen Unternehmensperformance auf die ökonomische Dimension feststellen können. Aufgrund dieser Erkenntnisse lassen sich zwei abschließende Hypothesen aufstellen:

> **Hypothese 9a:** *Die ökologische Unternehmensperformance hat einen positiven Einfluss auf die ökonomische Unternehmensperformance.*

> **Hypothese 9b:** *Die soziale Unternehmensperformance hat einen positiven Einfluss auf die ökonomische Unternehmensperformance.*

2.4.2. Theoretisches Forschungsmodell

Nachdem im vorherigen Abschnitt die Hypothesen zum Zusammenhang von SSCM Praktiken und der Unternehmensperformance im Sinne der TBL aus der Literatur hergeleitet wurden, werden diese nun zu einem theoretischen Forschungsmodell zusammengefügt. Das vorgeschlagene Forschungsmodell ist in Abbildung 3 dargestellt. Das Modell besteht aus jeweils vier Konstrukten von internen und externen SSCM Praktiken als exogene Variablen und drei weiteren Konstrukten der Unternehmensperformance als endogene Variablen. Alle hypothetischen Effekte zwischen den exogenen und endogenen Variablen sind positiv. Eine Besonderheit in diesem Modell sind die Konstrukte der ökologischen und sozialen Unternehmensperformance, denn sie können als Mediatorvariablen zwischen den SSCM Praktiken und der ökonomischen Unternehmensperformance fungieren. Dieses aus der Literatur hergeleitete Hypothesensystem auf Basis sachlogischer Beziehungen dient nun als Grundlage für die durchzuführende empirische Untersuchung. Wie sich die methodische Vorgehensweise gestaltet, wird im folgenden Kapitel erläutert.

3. Methodisches Vorgehen

Zur Beantwortung der zugrundeliegenden Forschungsfragen wird eine Befragung in Unternehmen nach der Survey Research Methodik durchgeführt. Im Folgenden wird deshalb in einem ersten Schritt die Survey Research Methodik als quantitatives Forschungsdesign und dessen Vorgehensweise theoretisch dargelegt. Im Anschluss wird auf die Zielgruppe sowie die Entwicklung des Fragebogens eingegangen und nachfolgend die Durchführung der Befragung und die endgültige Stichprobe beschrieben. Da die Auswertung der Befragung anhand einer PLS-SGM erfolgt, wird diese in Teilkapitel 3.5 erläutert. Abschließend wird das Datenmaterial auf Non-Response und Common-Method Bias überprüft.

3.1. Survey Research

Survey Research ist vor allem in den Sozialwissenschaften eine der am meist genutzten Forschungsmethoden (Hackett, 1981, S. 599). Aber auch in den Wirtschaftswissenschaften ist Survey Research ein beliebtes Forschungsinstrument (Faulbaum, 2019, S. XV). McGraw und Watson (1976, S. 343[5], zitiert nach Hackett, 1981, S. 600) definieren Survey Research als „a method of collecting standardized information by interviewing a sample representative of some population." Aufgrund zeitlicher und finanzieller Restriktionen ist es bei großen Zielpopulationen oftmals nicht möglich, eine Totalerhebung durchzuführen. Deshalb wird aus einer zuvor definierten Zielpopulation eine repräsentative Stichprobe ausgewählt, mit der wiederum Rückschlüsse auf die Zielpopulation gezogen werden sollen (Faulbaum, 2019, S. 3–4; Fowler, 1988, S. 9).

Mithilfe von Surveys (Befragungen) sollen Meinungen, Fakten und Einstellungen in Bezug auf den Forschungsgegenstand gesammelt werden. Der Hauptzweck einer Befragung kann deskriptiver, explanativer oder explorativer Art sein. Deskriptive Befragungen werden verwendet, um eine Population z. B. anhand demografischer Informationen, wie Alter, Geschlecht oder Herkunft zu beschreiben. Wenn Gründe für die Existenz bestimmter Fakten und Meinungen für eine Forschung von Interesse sind, handelt es sich um eine explanative Befragung. Explorative Befragungen werden durchgeführt, wenn bisher wenig über eine Population bekannt ist oder weitere Informationen zu Forschungsvariablen benötigt werden, bevor eine umfangreichere Forschung gestartet werden kann (Hackett, 1981, S. 600).

Hackett (1981, S. 602) schlägt ein sechsstufiges Phasenmodell zur Anwendung der Survey Research Methodik vor, welches in Abbildung 4 dargestellt wird. Die erste Phase beinhaltet die Problemdefinition des Forschungsprojektes. Hier stellt sich die Frage, ob Survey Research die geeignete Erhebungsart für die zugrundeliegende Fragestellung ist und welches Ziel damit erreicht werden soll (Hackett, 1981, S. 602). In der zweiten Phase wird das Survey Design ausgewählt. Hier geht es darum, festzulegen, welche Methoden am besten geeignet sind, um die erforderlichen Informationen zur Beantwortung der Forschungsfragen zu sammeln. Es wird grundsätzlich zwischen zwei Basisdesigns unterschieden. Während bei Querschnittsstudien Informationen zu einer Population zu einem bestimmten Zeitpunkt oder in einer bestimmten Zeitspanne gesammelt werden, werden bei Längsschnittstudien die Daten an zwei oder mehreren Zeitpunkten erhoben, um Veränderungen im Zeitverlauf feststellen zu können (Faulbaum, 2019, S. 48).

In der dritten Phase wird die Auswahl der Stichprobe vorgenommen. Es kann zwischen einer zufälligen und einer nicht-zufälligen Stichprobenauswahl unterschieden werden. Bei einer Zufallsstichprobe ist eine vollständige Liste der Zielpopulation erforderlich, aus welcher dann zufällig anhand verschiedener Auswahlverfahren die Stichprobe ausgewählt wird (Fowler, 1988, S. 20). Bei einer nicht-zufälligen Stichprobe wird anhand verschiedener Kriterien eine Stichprobe aus der Zielpopulation gezogen (Roy, Acharya & Roy, 2016,

[5] McGraw und Watson (1976): Political and social inquiry, Wiley New York, 1976.

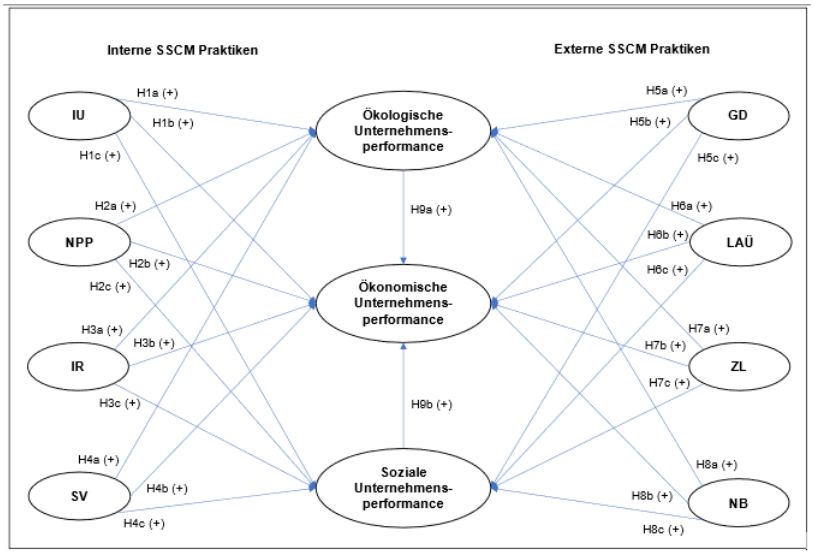


Abbildung 3: Theoretisches Forschungsmodell

Quelle: eigene Darstellung

Abbildung 4: Phasen des Survey Research Prozesses

Quelle: eigene Darstellung, nach (Hackett, 1981, S. 602)

S. 4). In der vierten Phase von Survey Research wird der Fragebogen entwickelt. Es stellt sich die Frage, welche Informationen gesammelt werden müssen, um die Forschungsfragen zu beantworten und welche Struktur (strukturiert oder unstrukturiert) der Fragebogen annehmen soll (Fowler, 1988, S. 99–106). Die fünfte Phase beinhaltet die Frage nach der Art der Durchführung der Survey, um bestmöglich die gewünschten Informationen sammeln zu können. Die gängigsten Methoden zur Datensammlung sind z. B. persönliche Interviews, Telefoninterviews, E-Mail-Surveys oder auch webbasierte Surveys (Faulbaum, 2019, S. 12). In der abschließenden sechsten Phase müssen die gesammelten Daten hinsichtlich der Forschungsfragen analysiert und interpretiert werden. Die Auswahl der Analysemethode sollte jedoch schon zu Beginn der Forschung festgelegt werden, da diese einen direkten Einfluss auf die Entwicklung von Fragebögen hat (Hackett, 1981, S. 603). Anhand der Struktur des Phasenmodells nach Hackett (1981, S. 602) wird nun weiter vorgegangen. Wie in der Einleitung schon erwähnt, soll die zugrundeliegende Forschungsfrage empirisch beantwortet werden, indem eine Querschnittsstudie mit Primärdatenanalyse durchgeführt wird. Im nächsten Teilkapitel wird deshalb gemäß der dritten Phase die Auswahl der Stichprobe erläutert.

3.2. Auswahl der Stichprobe

Ein signifikanter Anteil der globalen THG Emissionen wird durch industrielle Prozesse verursacht, welche somit ein hohes Minderungs- und Anpassungspotential aufweisen (Charkovska et al., 2019, S. 907). In Deutschland war der Industriesektor im Jahre 2016 für ca. 20,7 % der gesamten THG Emissionen verantwortlich (BMU, 2018, S. 34). Insbesondere die Supply Chains von Industrieunternehmen machen einen großen Anteil der THG Emissionen aus. Im Durchschnitt sind ca. 75 % des CO_2-Fußabdrucks eines Industriesektors auf dessen Supply Chain zurückzuführen (Huang et al., 2009, S. 8509). Während der Literaturrecherche konn-

ten nur wenige Publikationen ausfindig gemacht werden, die sich mit SSCM in deutschen Industrieunternehmen auseinandersetzten (z. B. Thun & Müller, 2009; Hansen et al., 2011; Hunke & Prause, 2014). Weiterhin wurde in diesen Studien kein Zusammenhang zwischen SSCM und der Unternehmensperformance untersucht. Aus diesen Gründen ist das herstellende und verarbeitende Gewerbe in Deutschland interessant für die durchzuführende Befragung.

Die Auswahl der Stichprobe erfolgt mit Hilfe der Amadeus Datenbank über vergleichbare Finanzdaten für börsennotierte und private Unternehmen in Europa. Über die Amadeus Datenbank kann auf detaillierte Informationen zu ca. 21 Millionen Unternehmen in ganz Europa zugegriffen werden (Bureau van Dijk 2020). Es werden die 500 größten Unternehmen in Deutschland nach Umsatz und des Abschnitts C (Verarbeitendes Gewerbe / Herstellung von Waren) der „statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft“, der sogenannte NACE Rev. 2 Code (franz. Nomenclature statistique des activités économiques dans la Communauté européenne), für die Befragung ausgewählt. Der NACE Rev. 2 Code ist ein System der EU zur Klassifizierung von Wirtschaftszweigen auf Basis der „International Standard Industrial Classification of all Economic Activities“ (ISIC) (Europäische Kommission, 2008). Der Abschnitt C wird in 13 Untergruppen (CA bis CM) und insgesamt 24 Abteilungen (10 bis 33) nach Branchen und Geschäftsfeldern aufgegliedert (Europäische Kommission, 2008, S. 45). Aufgrund der Übersichtlichkeit beschränkt sich die Aufgliederung der Stichprobe auf die 13 Untergruppen des Abschnitts C. Es handelt sich hierbei demnach um eine nicht-zufällige Stichprobe (Roy et al., 2016, S. 4).

Die Verteilung der 500 größten Unternehmen Deutschlands nach Umsatz, des Abschnitts C und nach den 13 Untergruppen kann der Tabelle 3 entnommen werden. 66 (13,2 %) Unternehmen sind der Gruppe Metallerzeugung und -bearbeitung, Herstellung von Metallerzeugnissen (CH), 65 (13 %) Unternehmen sind der Gruppe Maschinenbau (CK) und 60 (12 %) Unternehmen der Gruppe Herstellung von Nahrungs- und Genussmitteln, Getränken und Tabakerzeugnissen usw. zuzuordnen. Den kleinsten Anteil machen Unternehmen aus dem Bereich Kokerei und Mineralölverarbeitung aus (2 %). Diese Stichprobe gilt als Grundlage für die durchzuführende Befragung.

3.3. Entwicklung des Fragebogens

Für die Erhebung der Primärdaten wurde ein strukturierter, webbasierter Fragebogen mit programmierten Fragen auf der Umfrageplattform Qualtrics erstellt. Um valide Indikatoren für die zuvor festgelegten Konstrukte (vgl. Abbildung 3) zu bestimmen, wurden ausschließlich Konstrukte und deren Indikatoren verwendet, die sich bereits in einschlägiger wissenschaftlicher Literatur etabliert haben. Ein Pretest des Fragebogens wurde deshalb nicht durchgeführt. Jedes Konstrukt wurde anhand von Multi-Item-Skalen auf einer 5-Punkte-Likert-Skala von *[1] stimme überhaupt nicht zu* bis *[5] stimme voll zu* gemessen. Der Fragebogen besteht aus drei Abschnitten mit insgesamt 18 Fragen. Im ersten Abschnitt sollten die Befragten angeben, inwieweit sie der Ansicht sind, dass SSCM Praktiken in deren Unternehmen umgesetzt werden. Im zweiten Abschnitt sollten die Befragten angeben, inwieweit sie der Ansicht sind, dass deren Unternehmen in den vergangenen drei Jahren die ökologische, ökonomische und soziale Unternehmensperformance durch den Einsatz von SSCM Praktiken beeinflusst haben. Im dritten Abschnitt wurden typische demografische Daten zu Unternehmensgröße, Branche und zur befragten Person erhoben. Die Unternehmensgröße, gemessen am Jahresumsatz des abgelaufenen Geschäftsjahres dient dabei als für eine in der Unternehmensforschung übliche Kontrollvariable. Der vollständige Fragebogen kann der Tabelle A.1 im Anhang A entnommen werden.

Vier interne SSCM Praktiken wurden in Kapitel 2.4.1 identifiziert. Das Konstrukt IU wurde anhand von sechs Indikatoren gemessen, welche aus Zhu et al. (2005, S. 460), Rao und Holt (2005, S. 914), Zhu, Geng und Lai (2010, S. 1330) und Guang Shi, Lenny Koh, Baldwin und Cucchiella (2012, S. 57) entnommen wurden. NPP wurde anhand von fünf Indikatoren gemessen, welche sich nach Rao und Holt (2005, S. 915), Zhu et al. (2005, S. 460) sowie Carter, Kale und Grimm (2000, S. 226) richten. Die drei Indikatoren zur Messung des Konstrukts IR wurden von Zhu und Sarkis (2004, S. 276) sowie Zhu et al. (2010, S. 1330) übernommen. Das vierte interne Konstrukt SV wurde anhand von sechs Indikatoren gemessen, welche von Carter und Jennings (2002, S. 153) sowie Zhu et al. (2016, S. 420) entnommen wurden.

Es konnten weitere vier externe SSCM Praktiken in der Literatur identifiziert werden. GD wurde anhand von drei Indikatoren gemessen, wobei diese von Rao und Holt (2005, S. 915) sowie Ageron et al. (2012, S. 179–180) entnommen wurden. Die fünf Indikatoren des Konstrukts LAÜ stammen von Carter et al. (2000, S. 226), Krause et al. (2000, S. 42), Klassen und Vachon (2003, S. 349–350) sowie Zhu et al. (2008, S. 586). Das Konstrukt ZL wurde mithilfe von vier Indikatoren gemessen. Diese wurden von Rao und Holt (2005, S. 916) sowie Zhu et al. (2005, S. 460) entnommen. Das vierte externe Konstrukt NB, welches ebenfalls anhand von vier Indikatoren gemessen wurde, ist anhand der Fragebögen von Carter, Ellram und Ready (1998, S. 38), Zhu et al. (2007, S. 1051) sowie Zhu et al. (2010, S. 1329) zusammengestellt worden.

Gemäß der TBL wurden drei endogene Konstrukte der Unternehmensperformance identifiziert. Die ökologische Unternehmensperformance wurde anhand von sechs Indikatoren gemessen, welche von Zhu und Sarkis (2004, S. 277) sowie Zailani et al. (2012, S. 339) übernommen wurden. Fünf Indikatoren zur Messung der ökonomischen Unternehmensperformance stammen von Carter et al. (2000, S. 223), Zhu und Sarkis (2004, S. 277), Rao und Holt (2005, S. 915) sowie Paulraj (2011, S. 35–36). Die sechs Indikatoren der sozialen Unternehmensperformance sind aus Paulraj (2011, S. 37), Zailani et al. (2012, S. 339) sowie Das (2017, S. 1360) entnommen worden.

Tabelle 3: Datengrundlage für die Befragung

NACE Rev. 2		Anzahl (%)
CA	Herstellung von Nahrungs- und Genussmitteln, Getränken und Tabakerzeugnissen	60 (12,0 %)
CB	Herstellung von Textilien, Bekleidung, Leder, Lederwaren und Schuhen	13 (3,0 %)
CC	Herstellung von Holzwaren, Papier, Pappe und Waren daraus, Herstellung von Druckerzeugnissen	28 (5,6 %)
CD	Kokerei und Mineralölverarbeitung	9 (2,0 %)
CE	Herstellung von chemischen Erzeugnissen	39 (7,8 %)
CF	Herstellung von pharmazeutischen Erzeugnissen	28 (5,6 %)
CG	Herstellung von Gummi- und Kunststoffwaren sowie von Glas und Glaswaren, Keramik, Verarbeitung von Steinen und Erden	42 (8,4 %)
CH	Metallerzeugung und -bearbeitung, Herstellung von Metallerzeugnissen	66 (13,2 %)
CI	Herstellung von Datenverarbeitungsgeräten, elektronischen und optischen Erzeugnissen	37 (7,0 %)
CJ	Herstellung von elektrischen Ausrüstungen	39 (7,8 %)
CK	Maschinenbau	65 (13,0 %)
CL	Fahrzeugbau	50 (10,0 %)
CM	Sonstige Herstellung von Waren, Reparatur und Installation von Maschinen und Ausrüstungen	24 (5,0 %)
	Summe	**500** (100 %)

Quelle: eigene Darstellung, Abruf aus Amadeus Datenbank am 11.05.2020

3.4. Durchführung der Befragung, deskriptive Statistiken und Prüfung der Daten

Wie im vorangegangen Teilkapitel beschrieben, wurde die Online-Befragung mithilfe der Umfrageplattform Qualtrics durchgeführt. Über die Startseite der Umfrageplattform wurde den potenziellen Befragten ein Informationsblatt zur Verfügung gestellt, damit diese einen groben Überblick über das Forschungsvorhaben der vorliegenden Arbeit bekommen konnten (vgl. Abbildung A.1 im Anhang A). Die potenziellen Befragten sollten möglichst aus den Funktionsbereichen SCM, Einkauf, Operations, Logistik oder Nachhaltigkeitsmanagement stammen und eine Position im mittleren und höheren Management einnehmen, um sicherzustellen, dass die Befragten über weitreichende Kenntnisse über die SSCM Praktiken und die Supply Chain ihres Unternehmens verfügen (sog. Key-Informanten) (Campbell, 1955, S. 339; Malhotra & Grover, 1998, S. 423). Zur Überprüfung der Key-Informanten diente die Erhebung personenbezogener Daten bzgl. der Position im Unternehmen und der Länge der Beschäftigung in dieser Position (vgl. Kapitel 3.3 und Tabelle A.1 im Anhang A).

Die Querschnittsstudie, die über insgesamt neun Wochen durchgeführt wurde, erfolgte in zwei Phasen. In der ersten Phase, die am 01. Juni 2020 begann und am 14. Juni 2020 endete, wurde den 500 Unternehmen (vgl. Tabelle 3) eine Einladung zur Befragung inklusive Zugangslink zur Umfrageplattform per E-Mail zugesandt. In der zweiten Phase, die am 22. Juni 2020 begann und am 24. Juli 2020 endete, wurden 157 der Unternehmen per Telefon noch einmal an die Befragung erinnert. Am 31. Juli 2020 endete die Befragung.

Insgesamt wurde der Umfragelink 85-mal aufgerufen, was einer Rücklaufquote von 17,00 % entsprechen würde. Jedoch haben 24 Befragte den Fragebogen nicht vollständig beantwortet. Da nur vollständig ausgefüllte Fragebögen in die Auswertung einfließen, wurden die unvollständigen Fragebögen für die Auswertung ausgeschlossen. Demzufolge sind 61 Fragebögen vollständig beantwortet worden, was zu einer Rücklaufquote von insgesamt 12,20 % führt.[6] Davon wurden 59 Fragebögen in deutscher, zwei Fragebögen in englischer Sprache ausgefüllt. Nach Start der Umfrage über den Link konnten die Befragten den Fragebogen innerhalb von zwei Wochen bearbeiten. Die durchschnittliche Bearbeitungszeit der Befragten, die den Fragebogen am selben Tag starteten und beendeten, betrug ca. 22 Minuten. Gemäß Tabelle 4 stammen die meisten Antworten aus Unternehmen der Metallerzeugung und -bearbeitung, Herstellung von Metallerzeugnissen (16,39 %), gefolgt von Unternehmen aus dem Bereich sonstige Herstellung von Waren, Reparatur und Installation von Maschinen und Ausrüstungen (14,75 %) und Unternehmen aus dem Bereich Herstellung von Nahrungs- und Genussmitteln, Getränken und Tabakerzeugnissen sowie Maschinenbau (jeweils 11,48 %).

Der größte Anteil der Befragten ist in einem Unternehmen mit 10.000 oder mehr Mitarbeiter*innen beschäftigt (34,43 %), gefolgt von Unternehmen mit 1.000 bis 4.999

[6] Drei Fragebögen wurden von persönlichen Kontakten in Unternehmen beantwortet, was einen Anteil von ca. 5 % ausmacht. Eine Verzerrung der Daten konnte durch einen negativen Test auf Common-Method-Bias ausgeschlossen werden (vgl. Kapitel 3.6).

Tabelle 4: Finale Stichprobe

NACE Rev. 2		Anzahl (%)
CA	Herstellung von Nahrungs- und Genussmitteln, Getränken und Tabakerzeugnissen	7 (11,48 %)
CB	Herstellung von Textilien, Bekleidung, Leder, Lederwaren und Schuhen	1 (1,64 %)
CC	Herstellung von Holzwaren, Papier, Pappe und Waren daraus, Herstellung von Druckerzeugnissen	6 (9,48 %)
CD	Kokerei und Mineralölverarbeitung	0 (0,00 %)
CE	Herstellung von chemischen Erzeugnissen	5 (8,20 %)
CF	Herstellung von pharmazeutischen Erzeugnissen	3 (4,92 %)
CG	Herstellung von Gummi- und Kunststoffwaren sowie von Glas und Glaswaren, Keramik, Verarbeitung von Steinen und Erden	2 (3,28 %)
CH	Metallerzeugung und -bearbeitung, Herstellung von Metallerzeugnissen [a]	10 (16,39 %)
CI	Herstellung von Datenverarbeitungsgeräten, elektronischen und optischen Erzeugnissen	3 (4,92 %)
CJ	Herstellung von elektrischen Ausrüstungen [a]	2 (3,28 %)
CK	Maschinenbau	7 (11,48 %)
CL	Fahrzeugbau	6 (9,84 %)
CM	Sonstige Herstellung von Waren, Reparatur und Installation von Maschinen und Ausrüstungen [a]	9 (14,75 %)
	Summe	**61 (100 %)**

Anmerkungen: [a] *Im Fragebogen konnten die Befragten im Freitextfeld „andere“ (vgl. Tabelle A.1 im Anhang A) eigene Branchenbezeichnungen angeben, welche dann anschließend gemäß des NACE Rev. 2 (Europäische Kommission, 2008, S. 67–73) den passenden Abteilungen zugeordnet wurden.*
Quelle: eigene Darstellung, eigene Erhebung

Tabelle 5: Demografische Daten der befragten Unternehmen

Vollzeitäquivalente Mitarbeiter*innen	Anzahl	%	Jahresumsatz in EUR	Anzahl	%
weniger als 250	2	3,28 %	bis 100 Mio.	4	6,56 %
250 bis 499	1	1,64 %	bis 1 Mrd.	20	32,79 %
500 bis 999	6	9,84 %	bis 5 Mrd.	22	36,07 %
1.000 bis 4.999	17	27,87 %	bis 10 Mrd.	4	6,56 %
5.000 bis 9.999	14	22,95 %	bis 50 Mrd.	7	11,48 %
10.000 oder mehr	21	34,43 %	mehr als 50 Mrd.	4	6,56 %
Summe	**61**	**100 %**	**Summe**	**61**	**100 %**

Quelle: eigene Darstellung, eigene Erhebung

(27,87 %) und 5.000 bis 9.999 Mitarbeiter*innen (22,95 %). Von den 61 Befragten sind 36,07 % in einem Unternehmen mit einem Jahresumsatz des abgelaufenen Geschäftsjahres von bis zu 5 Mrd. EUR, 32,79 % in einem Unternehmen mit bis zu 1 Mrd. EUR sowie 11,48 % in einem Unternehmen mit bis zu 50 Mrd. EUR beschäftigt (vgl. Tabelle 5).

Um qualitativ hochwertige Antworten zu erlangen, sollten die potenziellen Befragten möglichst Positionen in den Bereichen Supply Chain, Logistik, Operations, Einkauf oder Nachhaltigkeitsmanagement besetzen. Demnach haben 49,18 % der Befragten eine Position in den Bereichen Supply Chain, Logistik, Operations oder Einkauf und 22,95 % in den Bereichen Qualitäts- und Nachhaltigkeitsmanagement oder CSR. Ein Anteil von 47,54 % der Befragten ist bereits mehr als 5 Jahre in ihrer aktuellen Position, 42,62 % zwischen ein bis fünf Jahren und 9,84 % weniger als ein Jahr (vgl. Tabelle 6). Insgesamt sind 77,05 % der 61 Befragten in einer Führungsposition und 93,44 % haben einen akademischen Abschluss.

Die deskriptiven Statistiken der gewonnen Daten aus der

Tabelle 6: Funktionsbereich und Jahre in aktueller Position der Befragten

Funktionsbereich	Anzahl	(%)	Jahre in aktueller Position	Anzahl	(%)
Geschäftsleitung, Werksleitung	6	(9,84 %)	weniger als 1	6	(9,84 %)
SCM, Logistik, Operations, Einkauf	30	(49,18 %)	1 bis 5	26	(42,62 %)
Marketing, Vertrieb	6	(9,84 %)	mehr als 5	29	(47,54 %)
Controlling, Finanzen	2	(3,28 %)			
Qualitäts-, Nachhaltigkeitsmanagement, CSR	14	(22,95 %)			
andere [a]	3	(4,92 %)			
Summe	**61**	**(100 %)**	**Summe**	**61**	**(100 %)**

Anmerkungen: [a] *Die Kategorie „andere" setzt sich aus Unternehmenskommunikation (2x) und IT (1x) zusammen.*
Quelle: eigene Darstellung, eigene Erhebung

Befragung wurden mit IBM SPSS Statistics 26 berechnet. Diese können Tabelle B.1 im Anhang B entnommen werden. Damit eine hohe Datenqualität sichergestellt werden kann, sollte das aus der Befragung gewonnene Datenmaterial sorgfältig überprüft werden. Ein besonderes Interesse bei der Datenüberprüfung liegt auf fehlenden Werten, inkonsistenten Antworten, Antwortmustern und die Verteilung der Daten (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 48). Wie beschrieben, wurden alle unvollständig beantworteten Fragebögen aus der Stichprobe entfernt, weshalb keine fehlenden Werte mehr existieren. In Hinblick auf inkonsistente Antworten konnten nach Durchsicht der Daten keine Auffälligkeiten festgestellt werden. Weiterhin wurden die Antworten visuell auf Straight-Lining, Diagonal-Lining und alternierende Extremantworten überprüft (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 50), wobei auch hier keine Auffälligkeiten identifiziert wurden. Um Antwortmuster gänzlich ausschließen zu können, wird in Kapitel 3.6 der Test auf Common-Method-Bias hinzugezogen. Zur Überprüfung der Datenverteilung werden die Werte für die Schiefe und Kurtosis herangezogen. Allgemein deuten Werte für Schiefe und Kurtosis von nahe 0 auf eine Normalverteilung der Daten, Werte von größer als 1 bzw. kleiner als -1 auf nicht-normalverteilte Daten hin. Bei der Schiefe ist eine Spanne mit Werten von -2,80 bis 0,86 zu beobachten. Positive bzw. negative Werte für die Schiefe weisen darauf hin, dass die Verteilung der Daten rechts- bzw. linksschief ist. Insgesamt sind 24 Werte kleiner als -1. Zudem kann festgestellt werden, dass bis auf ZL_3 und ZL_4 alle Indikatoren einen negativen Wert für die Schiefe aufweisen und damit linksschief verteilt sind. Bei der Kurtosis ist eine Spanne von -1,31 bis 7,83 zu beobachten, wovon 14 Werte größer 1 und 6 Werte kleiner -1 sind (vgl. Tabelle B.1 im Anhang B). Positive bzw. negative Werte für die Kurtosis weisen darauf hin, dass die Verteilung der Daten zu spitz bzw. zu flach ist. Da die zur Analyse anzuwendende PLS-SGM als nicht-parametrisches statistisches Verfahren gegenüber nicht-normalverteilten Daten sehr robust ist (hierzu Kapitel 3.5), sollten die festgestellten Abweichungen zur Normalverteilung keine erheblichen Probleme für die weitere Analyse der Daten darstellen (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 52).

3.5. Partial Least Squares Strukturgleichungsmodellierung

Die SGM, im deutschen auch häufig als Kausalanalyse bezeichnet, wird in der Wirtschafts- und Sozialforschung mit großer Beliebtheit angewandt, denn als multivariate Analysemethode ermöglichen SGMs die gleichzeitige Analyse der Qualität mehrerer Konstrukte hinsichtlich ihrer Reliabilität und Validität (Henseler, 2005, S. 70). Konstrukte enthalten z. B. Informationen zu verschiedenen Individuen, Unternehmen oder Ereignissen. Diese Informationen können entweder durch Primärforschung gewonnen werden oder auf Sekundärdaten beruhen (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 2). Das Ziel der SGM ist die Überprüfung kausaler Abhängigkeiten zwischen Konstrukten anhand eines Datensatzes. Als hypothesenprüfendes Verfahren setzt es voraus, dass vor Anwendung sachlogische Überlegungen zu Beziehungen zwischen den Variablen hergestellt werden. Anhand dieses theoretisch hergeleiteten Hypothesensystems wird dann überprüft, ob die aufgestellten Beziehungen zwischen den Konstrukten mit dem empirisch gewonnenen Datenmaterial übereinstimmen und wie stark diese Beziehungen sind. Die SGM wird somit durch einen konfirmatorischen Charakter geprägt. Das Besondere an der SGM ist, dass auch Beziehungen zwischen latenten Konstrukten überprüft werden können (Backhaus, Erichson & Weiber, 2011, S. 65). Latente Konstrukte entziehen sich ihrer direkten empirischen Beobachtbarkeit, weshalb geeignete Messmodelle generiert werden müssen, um diese anhand von Indikatoren messbar zu machen (Weiber & Mühlhaus, 2010, S. 19). Besonders zwei Ansätze sind bei der SGM mit latenten Konstrukten von großer Bedeutung: Zum einen die Kovarianzstrukturanalyse,

welche auch als LISREL bezeichnet wird, zum anderen der varianzanalytische Ansatz, die Methode der kleinsten partiellen Quadrate (PLS) nach Wold (1975) (Henseler, 2005, S. 70; Weiber & Mühlhaus, 2010, S. 253).[7]

Zur Beantwortung der Forschungsfrage wird das theoretische Forschungsmodell (vgl. Abbildung 3) mit Hilfe der PLS-SGM überprüft, da es sowohl niedrige Anforderungen an das Datenmaterial als auch an das Forschungsmodell stellt (Ringle, Sarstedt & Straub, 2012, S. iv). In Bezug auf die Stichprobengröße bietet das Verfahren auch bei kleinen Stichproben generell eine hohe Teststärke (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 16). Einer Daumenregel von Barclay, Higgins und Thompson (1995, S. 292) zufolge, sollte die Stichprobengröße entweder so groß sein, wie „das 10-fache der höchsten Anzahl an formativen Indikatoren, die zur Messung eines einzelnen Konstrukts verwendet werden, oder das 10-fache der höchsten Anzahl an Strukturpfaden, die auf ein bestimmtes Konstrukt im Strukturmodell gerichtet sind“ (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 21). Gemäß dieser Daumenregel beträgt die höchste Anzahl der Strukturpfade, die im theoretischen Forschungsmodell der vorliegenden Arbeit (vgl. Abbildung 3) auf das Konstrukt „Ökonomische Unternehmensperformance“ zeigen, acht, wonach eine Stichprobengröße von mindestens N = 80 vorliegen sollte. Da die Stichprobengröße der durchgeführten Befragung bei N = 61 liegt, ist das theoretische Forschungsmodell jedoch zu komplex für das Datenmaterial. Um die Anzahl der Strukturpfade zu verringern, ist das Modell deshalb in zwei Submodelle nach internen (Submodell (SM) 1) und externen (SM2) SSCM Praktiken aufgeteilt worden, wodurch die höchste Anzahl an Strukturpfaden auf sechs verringert werden konnte (vgl. Abbildung C.1 und Abbildung C.2 im Anhang C). Weiterhin setzt die PLS-SGM keine strikten Vorgaben zur Verteilung der Daten voraus (Henseler, 2005, S. 70). Wenn z. B. die häufige Voraussetzung normalverteilter Daten bei traditionellen Analyseverfahren nicht erfüllt werden kann, ist die PLS-SGM zu favorisieren (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 23). Auch im Hinblick auf die Modelleigenschaften ist die PLS-SGM sehr flexibel. Es können sowohl reflektive als auch formative Messmodelle verarbeitet werden. Ferner können mit Hilfe der PLS-SGM hochkomplexe Modelle analysiert werden, solange die Daumenregel für die Stichprobengröße nicht verletzt wird (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 25; Bliemel, Eggert, Fassott & Henseler, 2005, S. 11).

PLS-Pfadmodelle bestehen aus zwei Elementen: Zum Ersten aus dem Strukturmodell (inneres Modell), welches die Beziehungen zwischen den latenten Konstrukten abbildet, und zum Zweiten aus dem Messmodell (äußeres Modell), das die Beziehungen zwischen den latenten Konstrukten und ihren Indikatoren darstellt (vgl. Anhang C). Besonders wichtig ist die Unterscheidung zwischen formativen und reflektiven Messmodellen. „In einem reflektiven Messmodell geben die Korrespondenzregeln eine Kausalitätsrichtung vor, die von der theoretischen Ebene zur Beobachtungsebene weist. Somit wird unterstellt, dass [...] [das latente Konstrukt] ihre zugeordneten Indikatoren verursacht“ (Fassott & Eggert, 2005, S. 36). Die Pfeile in einem reflektiven Messmodell zeigen deshalb von der latenten Variablen auf dessen Indikatoren. Ein umgekehrter Wirkungszusammenhang wird im formativ spezifizierten Messmodell widergespiegelt (Henseler, 2005, S. 71). Hier „geben die Korrespondenzregeln eine Kausalitätsrichtung vor, die von der Beobachtungsebene zur theoretischen Ebene weist. In diesem Modell verursachen die beobachtbaren Indikatoren [...] [das latente Konstrukt] [..]“ (Fassott & Eggert, 2005, S. 38). Die Pfeile in einem formativen Messmodell zeigen diesbezüglich von den Indikatoren auf deren latentes Konstrukt (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 43). In Bezug auf die latenten Konstrukte des theoretischen Forschungsmodells der vorliegenden Arbeit konnte nach den Entscheidungsregeln für formativ und reflektiv spezifizierte Messmodelle von Jarvis, MacKenzie und Podsakoff (2003, S. 203) festgestellt werden, dass alle latenten Konstrukte reflektiv spezifiziert sind (vgl. Tabelle D.1 im Anhang D).

3.6. Prüfung auf Non-Response Bias und Common-Method Bias

Besonders E-Mail und Online-Umfragen gelangen aufgrund ihres Non-Response Bias (Schweigeverzerrung) vermehrt in die Kritik. Ein Non-Response Bias entsteht, wenn die Befragten, die tatsächlich an der Umfrage teilgenommen haben andere Antworten geben als diejenigen, die nicht an der Befragung teilgenommen haben, weshalb beim Vorliegen eines Non-Response Bias nicht grundsätzlich auf die Grundgesamtheit geschlossen werden kann (Armstrong & Overton, 1977, S. 396). Nach Armstrong und Overton (1977, S. 397) kann angenommen werden, dass Personen, die erst aufgrund einer Erinnerung an der Befragung teilnehmen, sich ähnlich der nicht antwortenden Personen verhalten. Um zu überprüfen, dass der Datensatz keinem Non-Response Bias unterliegt, wurde dieser deshalb in zwei Gruppen aufgeteilt. Die erste Gruppe besteht aus 25 Befragten, die während der ersten Phase der Einladung zum Fragebogen per E-Mail folgten und diesen beantworteten. Die zweite Gruppe besteht aus 36 Befragten, die in der zweiten Phase den Fragebogen nach einer Erinnerung per Telefon beantworteten. Anhand eines zweiseitigen t-Tests auf Mittelwertgleichheit für unabhängige Stichproben wurde überprüft, ob statistisch signifikante Unterschiede zwischen den Antworten dieser beiden Gruppen existieren (vgl. Tabelle E.1 und Tabelle E.2 im Anhang E). Voraussetzung für den t-Test ist, dass Varianzhomogenität zwischen beiden Gruppen vorliegt. Anhand des Levene-Tests der Varianzgleichheit kann bestätigt werden, dass Varianzhomogenität vorliegt ($F(1,59) = 0{,}257$; $p = 0{,}614$; $N = 61$). Der t-Test zeigt, dass es keinen statistisch signifikanten Unterschied zwischen den beiden Gruppen gibt ($t(59) = 0{,}127$; $p = 0{,}899$; $N = 61$). Somit stellt der Non-Response Bias kein Problem für die weitere Datenanalyse dar. Der t-Test wurde mit IBM SPSS Statistics 26 durchgeführt.

[7] Auf eine genaue Beschreibung der Funktionsweise und einen Vergleich beider Ansätze wird an dieser Stelle verzichtet, da dies nicht den Zweck der vorliegenden Arbeit erfüllt.

Da die Primärdaten anhand einer Querschnittsstudie gesammelt wurden, könnte ein Common-Method Bias die Datenanalyse beeinträchtigen (Matzler, Strobl, Thurner & Füller, 2015, S. 124). Ein Common-Method Bias entsteht, wenn z. B. die Befragten gleichzeitig Fragen zu exogenen und endogenen Konstrukten beantworten (Podsakoff & Organ, 1986, S. 533–534) oder persönliche Kontakte den Fragebogen bearbeiten, was in der vorliegenden Befragung der Fall war. Um zu überprüfen, ob der Datensatz einem Common-Method Bias unterliegt, wurde ex post Harman's Single Factor Test durchgeführt. Hierbei wurden die Indikatoren aller Konstrukte der Befragung in eine explorative Faktoranalyse (EFA) einbezogen. Ein Common-Method Bias liegt vor, wenn „(a) a single factor will emerge from the factor analysis or (b) one general factor will account for the majority of the covariance among the measures" (Podsakoff, MacKenzie, Lee & Podsakoff, 2003, S. 889). Nach Extraktion der Faktoren nach der Hauptkomponentenmethode ergaben sich 13 verschiedene Faktoren mit Eigenwerten größer 1 (Kaiserkriterium nach Kaiser & Rice, 1974, S. 114), wobei der erste Faktor ca. 30,25 % der Gesamtvarianz erklärte (vgl. Tabelle E.3 im Anhang E). Anschließend wurde erneut eine EFA durchgeführt, bei der die Anzahl der zu extrahierenden Faktoren auf 1 festgelegt wurde. Der extrahierte Faktor erklärt mit ebenfalls nur ca. 30,25 % nicht die Mehrheit der Gesamtvarianz (vgl. Tabelle E.4 im Anhang E). Demnach kann davon ausgegangen werden, dass ein Common-Method Bias nach Harman's Single Factor Test kein Problem für die anschließende Datenanalyse zu sein scheint. Dieser Test wurde mit IBM SPSS Statistics 26 durchgeführt.

Da Harman's Single Factor Test aber aufgrund seiner geringen Aussagekraft in der Literatur häufig kritisiert wird (z. B. Podsakoff et al., 2003, S. 890), wurde zusätzlich der Full Collinearity Test nach Kock (2015, S. 7) durchgeführt. Mithilfe dieses Tests werden die Kollinearitäten zwischen allen latenten Konstrukten in beiden Submodellen überprüft. Zur Beurteilung der Kollinearitäten wird der sogenannte Varianzinflationsfaktor (VIF) herangezogen. Ein VIF-Wert von kleiner als 3,3, mindestens aber kleiner als 5, deutet darauf hin, dass kein Common-Method-Bias im Modell vorliegt (Kock, 2015, S. 7; Hair, Ringle & Sarstedt, 2011, S. 145). Der Full Collinearity Test wurde mithilfe der Analysesoftware SmartPLS 3.3.2 (Ringle, Wende & Becker, 2020) durchgeführt. Weder in SM1 noch in SM2 wurde ein kritischer VIF-Wert von 5 oder höher zwischen den latenten Konstrukten festgestellt (vgl. Tabelle E.5 für SM1 und Tabelle E.6 für SM2 im Anhang E). Der höchste VIF-Wert in SM1 beträgt 2,390 für die Kollinearität zwischen SozUP und SV und ist somit deutlich unter dem Schwellenwert von 3,3. In SM2 beträgt der höchste VIF-Wert 3,468 für die Kollinearität zwischen GD und ÖkolUP und liegt knapp über dem Schwellenwert von 3,3, aber deutlich unter dem Wert von 5. Somit kann auch nach dem Full Collinearity Test das Vorliegen eines Common-Method-Bias ausgeschlossen werden.

4. Datenanalyse und Ergebnisse

Nachdem im vorherigen Kapitel die zugrundeliegende Stichprobe der Befragung und die Grundlagen der Methodik vorgestellt wurden, werden in diesem Kapitel die gewonnenen Primärdaten analysiert und die Ergebnisse dargelegt. Obwohl die PLS-SGM die Verknüpfung von Indikatoren mit ihren latenten Konstrukten und die Beziehungen zwischen den latenten Konstrukten gleichzeitig schätzt, wird ein PLS-Pfadmodell üblicherweise in zwei aufeinanderfolgenden Schritten analysiert. Im ersten Schritt werden die Messmodelle der beiden theoretischen Submodelle auf Reliabilität und Validität geprüft. Daraufhin werden im zweiten Schritt die Strukturmodelle evaluiert. Dieses Vorgehen stellt sicher, dass das Modell über zuverlässige und gültige Konstruktmessungen verfügt, bevor Schlussfolgerungen über die Beziehungen zwischen den latenten Konstrukten hergestellt werden können (Hulland, 1999, S. 198). Die PLS-SGM wurde mit der Analysesoftware SmartPLS 3.3.2 von Ringle et al. (2020) durchgeführt.

4.1. Evaluation der Messmodelle

Zur Evaluation reflektiver Messmodelle bei einer PLS-SGM existieren verschiedene etablierte Evaluationskriterien. Die wichtigsten Evaluationskriterien sind die Interne-Konsistenz-Reliabilität sowie die Konvergenz- und Diskriminanzvalidität (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 90). Reliabilität drückt die Genauigkeit von Messinstrumenten aus. Validität beschreibt „das Ausmaß, mit dem ein Messinstrument auch das misst, was es messen sollte" (Weiber & Mühlhaus, 2010, S. 103). Die Überprüfung der Internen-Konsistenz-Reliabilität erfolgt anhand Cronbachs Alpha (CA) und der Composite-Reliabilität (CR). Die Konvergenzvalidität wird anhand der Faktorladungen, der Indikatorreliabilität sowie der durchschnittlich erfassten Varianz (AVE) überprüft. Abschließend werden zur Überprüfung der Diskriminanzvalidität die Kreuzladungen, das Fornell-Larcker (FL)-Kriterium sowie das Heterot-Monotrait (HTMT)-Verhältnis herangezogen (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 91). Bevor mit der Evaluation der Messmodelle begonnen werden konnte, galt es zu überprüfen, ob der PLS-Algorithmus konvergiert ist. Dies ist der Fall, wenn die maximale Anzahl an Iterationen nicht erreicht wurde (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 106–107). Für die Berechnung von SM1 und SM2 wurde eine maximale Anzahl an Iterationen von 1.000 vorgegeben. SM1 konvergierte nach 14, SM2 nach 8 Iterationen, wonach eine stabile Lösung erzeugt werden konnte. Die Einstellungen für den PLS-Algorithmus können der Abbildung F.1 im Anhang F entnommen werden.

Zunächst wurden die Faktorladungen der Indikatoren auf ihren latenten Konstrukten überprüft. Allgemein gilt, wenn die Faktorladungen der Indikatoren auf ihren latenten Konstrukten einen Wert größer als 0,70[8] aufweisen und

[8] In der ersten Iteration werden alle Indikatoren beibehalten, bei denen

statistisch signifikant sind, dass das Messmodell eine gute Konvergenzvalidität hervorbringt (Peng & Lai, 2012, S. 471). Die Faktorladungen aller Indikatoren können Tabelle G.1 (für SM1) und Tabelle G.2 (für SM2) im Anhang G entnommen werden. Anhand der genannten Kriterien wurden bei SM1 nach der ersten Iteration acht Indikatoren, nach der zweiten Iteration ein Indikator eliminiert (vgl. Tabelle G.1). Bei SM2 wurden nach der ersten Iteration vier Indikatoren, nach der zweiten Iteration ein Indikator eliminiert (vgl. Tabelle G.2). Da die PLS-SGM wegen fehlender Verteilungsannahmen nicht für parametrische Signifikanztests geeignet ist, wurde zur Überprüfung der statistischen Signifikanz der Faktorladungen das sogenannte Bootstrapping-Verfahren angewandt (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 103). „Beim Bootstrapping werden aus einem empirischen Datensatz wiederholt Stichproben (b = 1, 2, ..., B) einer festgelegten Größe (n*) mit Zurücklegen gezogen und mit deren Hilfe Teststatistiken berechnet“ (Weiber & Mühlhaus, 2010, S. 256). Die Einstellungen zur Durchführung des Bootstrapping-Verfahrens mit SmartPLS 3.3.2 können Abbildung F.2 im Anhang F entnommen werden. Nach Durchführung des Bootstrapping-Verfahrens konnte festgestellt werden, dass alle Faktorladungen statistische Signifikanz im 90 % Bias-corrected and accelerated (Bca)-Konfidenzintervall aufweisen (vgl. Tabelle G.1 und Tabelle G.2).

Zur Überprüfung der Internen-Konsistenz-Reliabilität wurde zunächst CA herangezogen. Nunnally und Bernstein (1994, S. 252[9], zitiert nach Weiber & Mühlhaus, 2010, S. 110) schlagen einen CA-Wert von mindestens 0,70 für ein latentes Konstrukt vor. Je näher der CA-Wert an 1 liegt, desto höher ist die Interne-Konsistenz-Reliabilität (Weiber & Mühlhaus, 2010, S. 110). Für SM1 und SM2 konnten nach der dritten Iteration bei allen latenten Konstrukten CA-Werte von über 0,70 festgestellt werden, was ein Hinweis für eine hohe Interne-Konsistenz-Reliabilität der latenten Konstrukte ist. Sowohl in SM1 als auch in SM2 weist das latente Konstrukt ÖkonUP mit 0,917 den höchsten CA-Wert auf, was für ein gutes Ergebnis spricht (vgl. Tabelle 7). Da CA als eher konservatives Maß aber dazu tendiert, die Interne-Konsistenz-Reliabilität zu unterschätzen, wurde als zweite Prüfgröße die CR herangezogen. Diese wird zwischen 0 und 1 definiert, wobei auch hier höhere Werte eine höhere Reliabilität indizieren. CR-Werte zwischen 0,60 und 0,95 gelten in der Literatur als zufriedenstellend (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 97). Sowohl die CR-Werte der latenten Konstrukte von SM1 als auch die von SM2 liegen innerhalb dieses Bereichs, weshalb auch hier von einer hohen Internen-Konsistenz-Reliabilität beider Submodelle ausgegangen werden kann (vgl. Tabelle 7).

Mit der zu Beginn dieses Kapitels durchgeführten Evaluation der Faktorladungen der einzelnen Indikatoren auf ihren latenten Konstrukten ist der erste Schritt der Prüfung der Messmodelle auf Konvergenzvalidität auf Indikatorebene bereits abgeschlossen. Ein zweites Prüfkriterium zur Konvergenzvalidität auf Indikatorebene ist die Indikatorreliabilität. Diese „gibt den Anteil der Varianz eines Indikators an, der durch das Konstrukt erklärt wird“ (Weiber & Mühlhaus, 2010, S. 122). Die Indikatorreliabilität entspricht dem Quadranten der jeweiligen Faktorladungen. In der Literatur wird ein Schwellenwert für eine angemessene Indikatorreliabilität von mindestens 0,40 empfohlen (Weiber & Mühlhaus, 2010, S. 122–124). Gemäß Tabelle H.1 im Anhang H betragen die Werte für die Indikatorreliabilität für SM1 0,377 bis 0,908 und für SM2 0,434 bis 0,906. Obwohl die Indikatorreliabilität für ÖkolUP_1 in SM1 mit 0,377 knapp unter dem Schwellenwert von 0,40 liegt, wird dieser Indikator analog zur Faktorladung aufgrund von Inhaltsvalidität beibehalten. Ein Gütekriterium zur Prüfung der Konvergenzvalidität auf Konstruktebene ist die AVE, welche „als der Mittelwert der quadrierten Ladungen aller mit dem [latenten] Konstrukt zusammenhängenden Indikatoren [...] definiert“ wird (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 99). Die AVE beschreibt, „wie viel Prozent der Streuung des latenten Konstruktes über die Indikatoren durchschnittlich erklärt wird“ (Weiber & Mühlhaus, 2010, S. 123). Nach Fornell und Larcker (1981, S. 46) sollte ein latentes Konstrukt im Schnitt mindestens die Hälfte der Varianz seiner Indikatoren erklären, weshalb die Autor*innen einen AVE-Wert von mindestens 0,50 vorschlagen. Tabelle 7 zeigt, dass für jedes latente Konstrukt der AVE-Wert oberhalb dieser Schwelle liegt. Sowohl in SM1 als auch in SM2 hat das latente Konstrukt ÖkonUP mit Werten von 0,858 die höchste AVE. Insgesamt kann also angenommen werden, dass sowohl für SM1 als auch für SM2 die Interne-Konsistenz-Reliabilität und die Konvergenzvalidität gegeben ist.

Als methodische Ergänzung der Konvergenzvalidität gilt die Prüfung auf Diskriminanzvalidität (Hulland, 1999, S. 199), welche das Ausmaß beschreibt, „in dem ein Konstrukt sich tatsächlich von anderen Konstrukten entlang empirischer Standards unterscheidet“ (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 99). Im Kontext der PLS-SGM ist das Hauptkriterium für eine angemessene Diskriminanzvalidität, dass das latente Konstrukt mehr Varianz mit seinen eigenen Indikatoren teilt, als mit anderen latenten Konstrukten im gleichen Modell (Hulland, 1999, S. 199). Als erstes Gütekriterium wird in der Literatur die Überprüfung der Kreuzladungen der Indikatoren genannt. Dementsprechend sollte die Faktorladung eines Indikators auf seinem theoretisch zugeordneten latenten Konstrukt höher sein als jede Kreuzladung auf einem anderen latenten Konstrukt (Chin, 1998b, S. 321). Die Kreuzladungen der jeweiligen Indikatoren von SM1 und SM2 können Tabelle I.1 und Tabelle I.2 im Anhang I entnommen werden. In SM1 hat z. B. der Indikator IU_2 mit 0,857 die höchste Ladung mit seinem zugeordneten latenten Konstrukt und weist deutlich geringere Ladungen zu den anderen latenten Konstrukten auf (z. B. IU_2 auf NPP

die Faktorladung kaufmännisch gerundet mindestens 0,70 ergibt. Da der Wert von 0,70 als sehr konservativ gilt, werden in der zweiten und dritten Iteration alle Indikatoren mit einer Faktorladung von mindestens 0,60 beibehalten, um auch die Inhaltsvalidität der Konstrukte zu wahren (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 98).

[9] Nunnally und Bernstein (1994): Psycometric theory, 3. Aufl., McGraw-Hill New York, 1994.

Tabelle 7: Interne-Konsistenz-Reliabilität und Konvergenzvalidität

Latentes Konstrukt	Interne-Konsistenz-Reliabilität		Konvergenzvalidität		
	CA	CR	FL (Anz. d. Items)	Ind.-Rel. (Spanne)	AVE
	≥ 0,70	0,60 ≤ CR ≤ 0,95	≥ 0,70 (≥ 0,60)	≥ 0,40	≥ 0,50
Submodell 1 (interne SSCM Praktiken)					
IU	0,872	0,907	0,727 - 0,857 (5)	0,529 - 0,734	0,661
NPP	0,748	0,856	0,743 - 0,885 (3)	0,552 - 0,783	0,666
IR	0,712	0,810	0,653 - 0,883 (3)	0,426 - 0,780	0,591
SV	0,776	0,855	0,767 - 0,782 (4)	0,588 - 0,612	0,596
ÖkolUP	0,822	0,883	0,614 - 0,920 (4)	0,377 - 0,846	0,659
ÖkonUP	0,917	0,948	0,878 - 0,953 (3)	0,771 - 0,908	0,858
SozUP	0,866	0,899	0,724 - 0,835 (6)	0,524 - 0,697	0,598
Submodell 2 (externe SSCM Praktiken)					
GD	0,737	0,883	0,868 - 0,910 (2)	0,753 - 0,828	0,791
LAÜ	0,899	0,925	0,766 - 0,882 (5)	0,587 - 0,778	0,713
ZL	0,858	0,904	0,810 - 0,854 (4)	0,656 - 0,729	0,702
NB	0,801	0,884	0,794 - 0,911 (3)	0,630 - 0,830	0,718
ÖkolUP	0,831	0,881	0,659 - 0,889 (5)	0,434 - 0,790	0,599
ÖkonUP	0,917	0,948	0,878 - 0,952 (3)	0,771 - 0,906	0,858
SozUP	0,866	0,899	0,716 - 0,826 (6)	0,513 - 0,682	0,598

Anmerkungen: ***AVE*** *= durchschnittlich erfasste Varianz;* ***CA*** *= Cronbachs Alpha;* ***CR*** *= Composite-Reliabilität;* ***FL*** *= Faktorladung (Spanne).*
Quelle: eigene Darstellung, berechnet mit SmartPLS 3.3.2 [Ringle, Wende & Becker 2020]

mit 0,274). In SM2 hat z. B. der Indikator GD_3 mit 0,910 die höchste Ladung mit seinem zugeordneten latenten Konstrukt GD und weist deutlich geringere Ladungen zu anderen latenten Konstrukten auf (z. B. GD_3 auf LAÜ mit 0,637). Dies kann für alle anderen Indikatoren beider Submodelle so fortgeführt werden. Da alle Indikatoren auf ihren zugeordneten latenten Konstrukten am höchsten laden, bestätigt das Gütekriterium der Kreuzladungen die Diskriminanzvalidität der latenten Konstrukte.

Die zweite Möglichkeit zur Überprüfung der Diskriminanzvalidität der latenten Konstrukte ist das FL-Kriterium nach Fornell und Larcker (1981, S. 45–46). Hierbei wird die Quadratwurzel der AVE eines latenten Konstrukts mit der Korrelation desselben latenten Konstrukts zu allen anderen latenten Konstrukten verglichen (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 100). Nach Fornell und Larcker (1981, S. 46) wird von Diskriminanzvalidität ausgegangen, solange die Quadratwurzel der AVE eines latenten Konstrukts größer ist als die Korrelationen zu allen anderen latenten Konstrukten.

In Tabelle 8 sind auf der Diagonalen die Quadratwurzel der AVE der jeweiligen latenten Konstrukte dargestellt. Auf der Nichtdiagonalen befinden sich jeweils die Korrelationen zu den anderen latenten Konstrukten (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 110). In SM1 weist z. B. die Quadratwurzel der AVE vom latenten Konstrukt NPP einen Wert von 0,816 auf. Dieser Wert wird nun mit den Korrelationen zu allen anderen latenten Konstrukten in der gleichen Spalte bzw. Zeile verglichen. Demnach kann festgestellt werden, dass der Wert 0,816 größer ist als alle anderen Korrelationswerte (z. B. IU mit 0,433 oder IR mit 0,066). Analog weist in SM2 z. B. die Quadratwurzel der AVE vom latenten Konstrukt ÖkonUP einen Wert von 0,926 auf. Verglichen mit den Korrelationen kann auch hier festgestellt werden, dass 0,926 der größte Wert in der entsprechenden Zeile bzw. Spalte ist (z. B. im Vergleich zu GD mit 0,280). Insgesamt kann nach dem FL-Kriterium konstatiert werden, dass für alle latenten Konstrukte von SM1 und SM2 Diskriminanzvalidität gegeben ist.

Als drittes Kriterium zur Überprüfung der Diskriminanzvalidität schlagen Henseler (2005, S. 120) das HTMT-Verhältnis der Korrelationen vor, da die Leistungsfähigkeit der Kreuzladungen sowie des FL-Kriteriums nach neuesten Studien unzuverlässig scheint. Bei der Analyse der Kreuzladungen kann es zu Problemen führen, wenn zwei latente Konstrukte hoch miteinander korrelieren. Das FL-Kriterium weist zudem Schwächen auf, wenn sich die Faktorladungen der Indikatoren der jeweiligen latenten Konstrukte nur leicht voneinander unterscheiden (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 102). Das HTMT-Verhältnis ist das Verhältnis von „den Korrelationen zwischen den Indikatoren, die unterschiedliche Konstrukte messen [...], und den Korrelationen zwischen Indikatoren, die jeweils ihr eigenes Konstrukt messen“ (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 102). Nach Henseler, Ringle und Sarstedt (2015, S. 123) sollte das HTMT-Verhältnis einen Wert von 0,85, mindestens aber 0,90 nicht überschreiten, denn ein Wert, der zu nah an 1 liegt, deutet auf redundante latente

Tabelle 8: FL-Kriterium der Submodelle

SM1	IU	NPP	IR	SV	ÖkolUP	ÖkonUP	SozUP
IU	**0,813**						
NPP	0,433	**0,816**					
IR	-0,166	0,066	**0,769**				
SV	0,670	0,425	-0,088	**0,772**			
ÖkolUP	0,597	0,418	-0,015	0,532	**0,812**		
ÖkonUP	0,100	0,294	0,019	0,179	0,334	**0,926**	
SozUP	0,515	0,349	-0,059	0,490	0,661	0,539	**0,773**
SM2	**GD**	**LAÜ**	**ZL**	**NB**	**ÖkolUP**	**ÖkonUP**	**SozUP**
GD	**0,889**						
LAÜ	0,732	**0,844**					
ZL	0,705	0,735	**0,838**				
NB	0,741	0,614	0,627	**0,847**			
ÖkolUP	0,507	0,565	0,513	0,441	**0,774**		
ÖkonUP	0,280	0,220	0,344	0,251	0,347	**0,926**	
SozUP	0,550	0,553	0,509	0,389	0,645	0,540	**0,773**

Quelle: eigene Darstellung, berechnet mit SmartPLS 3.3.2 (Ringle et al., 2012)

Tabelle 9: HTMT-Verhältnisse der Submodelle

SM1	IU	NPP	IR	SV	ÖkolUP	ÖkonUP	SozUP
IU							
NPP	0,544 ^{s}						
IR	0,206 ^{s}	0,167 ^{s}					
SV	0,803 ^{s}	0,530 ^{s}	0,196 ^{s}				
ÖkolUP	0,661 ^{s}	0,544 ^{s}	0,161 ^{s}	0,638 ^{s}			
ÖkonUP	0,117 ^{s}	0,356 ^{s}	0,066 ^{s}	0,222 ^{s}	0,393 ^{s}		
SozUP	0,564 ^{s}	0,420 ^{s}	0,188 ^{s}	0,581 ^{s}	0,764 ^{s}	0,589 ^{s}	
SM2	**GD**	**LAÜ**	**ZL**	**NB**	**ÖkolUP**	**ÖkonUP**	**SozUP**
GD							
LAÜ	0,908 $^{n.s.}$						
ZL	0,882 $^{n.s.}$	0,841^{s}					
NB	0,968 $^{n.s.}$	0,720^{s}	0,763^{s}				
ÖkolUP	0,623 ^{s}	0,625^{s}	0,591^{s}	0,744^{s}			
ÖkonUP	0,326 ^{s}	0,231^{s}	0,382^{s}	0,301^{s}	0,402^{s}		
SozUP	0,672 ^{s}	0,614^{s}	0,571^{s}	0,463^{s}	0,744^{s}	0,589^{s}	

Anmerkungen: ^{s} *= signifikant im 90 % Bca-Konfidenzintervall;* $^{n.s.}$ *= nicht signifikant im 90 % Bca-Konfidenzintervall (vgl. Tabelle J.1 und Tabelle J.2 im Anhang J).*
Quelle: eigene Darstellung, berechnet mit SmartPLS 3.3.2 (Ringle et al., 2012)

Konstrukte und somit auf einen Mangel an Diskriminanzvalidität hin. In Tabelle 9 werden alle Kombinationen der latenten Konstrukte mit ihren HTMT-Werten dargestellt.

In SM1 kann festgestellt werden, dass alle HTMT-Werte unter dem eher konservativen Schwellenwert von 0,85 liegen (z. B. IU und IR mit 0,206), wodurch von Diskriminanzvalidität der latenten Konstrukte ausgegangen werden kann. In SM2 liegen die Werte für die Konstruktkombinationen GD und LAÜ (0,908) sowie GD und NB (0,968) über dem Wert von 0,90 und der Wert für GD und ZL mit 0,882 zwischen 0,85 und 0,90. Dies kann auf Probleme hinsichtlich der Diskriminanzvalidität dieser Konstruktkombinationen hinweisen. Um dies zu überprüfen, konnte mithilfe des HTMT-Kriteriums statistisch getestet werden, ob sich die HTMT-Statistik signifikant von 1 unterscheidet. Mittels Bootstrapping-Verfahren wurden für alle HTMT-Werte Bca-Konfidenzintervalle mit einer Vertrauenswahrscheinlichkeit von 90 % ermittelt. Enthält das Bca-Konfidenzintervall

den Wert 1 nicht, so kann angenommen werden, dass die beiden latenten Konstrukte einer Kombination empirisch unterschiedlich sind und vice versa (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 103–104). Die 90 % Bca-Konfidenzintervalle der HTMT-Werte können Tabelle J.1 (für SM1) und Tabelle J.2 (für SM2) im Anhang J entnommen werden. In SM1 unterscheiden sich z. B. die latenten Konstrukte der Kombination IU und IR statistisch signifikant von 1 (90 % KI = 0,091; 0,269). In SM2 unterscheiden sich, wie schon bei der Prüfung der HTMT-Werte vermutet wurde, die Konstruktkombinationen GD und LAÜ (90 % KI = 0,755; 1,056), GD und NB (90 % KI = 0,834; 1,117) sowie GD und ZL (90 % KI = 0,752; 1,004) nicht signifikant von 1, was auf einen Mangel an Diskriminanzvalidität hindeutet. Alle anderen Konstruktkombinationen von SM2 unterscheiden sich signifikant von 1. Da für die problematischen Konstruktkombinationen von SM2 eine Diskriminanzvalidität über die Prüfung der Kreuzladungen und das FL-Kriterium festgestellt werden konnte, bleiben diese für die anschließende Evaluation der PLS-Strukturmodelle dennoch erhalten.

4.2. Evaluation der Strukturmodelle

Nachdem für alle Messmodelle von SM1 und SM1 Reliabilität und Validität festgestellt werden konnte, kann nun mit der Evaluation der Strukturmodelle fortgefahren werden. Das Vorgehen richtet sich weitestgehend nach Hair, Hult, Ringle, Sarstedt, Richter und Hauff (2017, S. 165). Zuerst werden die Kollinearitäten der Strukturmodelle überprüft. Im zweiten Schritt werden die standardisierten Pfadkoeffizienten der Beziehungen zwischen den exogenen und endogenen latenten Konstrukten im Strukturmodell inklusive Relevanz und Signifikanz getestet. Hierbei wird auf die einzelnen Hypothesen von SM1 und SM2 eingegangen. Anschließend werden die Strukturmodelle auf mögliche Mediatoreffekte untersucht. Daraufhin werden die Bestimmtheitsmaße (R^2-Werte) der latent endogenen Konstrukte sowie die f^2-Effektstärken überprüft. Im fünften Schritt wird mithilfe der Blindfolding-Prozedur die Prognoserelevanz der Strukturmodelle (Q^2-Wert) bestimmt, auf dessen Grundlage die q^2-Effektstärken ermittelt und überprüft werden. Abschließend wird überprüft, ob die Kontrollvariable der Unternehmensgröße einen Einfluss auf die beiden Submodelle hat und ob es signifikante Unterschiede beim Grad der Implementierung zwischen internen und externen SSCM Praktiken gibt.

Prüfung der Kollinearitätsstatistik

Als Kriterium zur Prüfung der Kollinearität in den Strukturmodellen wird erneut der VIF herangezogen. Nach Hair et al. (2011, S. 145) sollte der VIF-Wert kleiner als 5 sein, um eine Kollinearität im Strukturmodell ausschließen zu können. Im Umkehrschluss deuten alle VIF-Werte über diesem Schwellenwert auf Kollinearität im Strukturmodell hin. Für SM1 wird die Kollinearität der folgenden Sets an Treiberkonstrukten geprüft: IU, NPP, IR und SV als Treiber von ÖkolUP, ÖkonUP und SozUP sowie ÖkolUP und SozUP als Treiber von ÖkonUP. Für SM2 werden analog die folgenden Sets geprüft: GD, LAÜ, ZL und NB als Treiber von ÖkolUP, ÖkonUP und SozUP sowie ÖkolUP und SozUP als Treiber von ÖkonUP. Wie Tabelle K.1 (für SM1) und Tabelle K.2 (für SM2) im Anhang K zeigen, befinden sich alle VIF-Werte deutlich unter dem Schwellenwert von 5 (z. B. IU und ÖkolUP mit VIF = 1,976), wodurch kein kritisches Maß an Kollinearität zwischen den Treiberkonstrukten in den Strukturmodellen festgestellt werden konnte.

Prüfung der Pfadkoeffizienten der Strukturpfade

Mithilfe des PLS-Algorithmus wurden die standardisierten Pfadkoeffizienten (β) geschätzt. Diese liegen normalerweise in einem Intervall von -1 bis 1, wobei ein Wert nahe 1 eine stark positive Beziehung repräsentiert (vice versa für negative Werte) und ein Wert nahe 0 eine eher schwache Beziehung darstellt (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 168). Nach Chin (1998a, S. xiii) gelten standardisierte Pfadkoeffizienten bereits ab einem Wert von 0,200 bzw. -0,200 als bedeutungsvoll. Um die 90 % Bca-Konfidenzintervalle zur Überprüfung der Signifikanz der standardisierten Pfadkoeffizienten zu erhalten, wurde erneut das Bootstrapping-Verfahren angewandt.

Nachfolgend werden im ersten Schritt die direkten standardisierten Pfadkoeffizienten von SM1 und SM1 systematisch nach den in Kapitel 2.4 aufgestellten Hypothesen dargelegt. Die Werte der standardisierten Pfadkoeffizienten mit ihren jeweiligen Signifikanzniveaus können Abbildung 5 (für SM1) und Abbildung 6 (für SM2) entnommen werden. Zusätzlich werden die p-Werte, t-Werte und 90 % Bca-Konfidenzintervalle in Tabelle L.1 (für SM1) und Tabelle L.2 (für SM2) im Anhang L dargestellt.

Für die Hypothesen H1a, H1b und H1c wurde ein positiver Einfluss von IU auf die ökologische (H1a), ökonomische (H1b) und die soziale (H1c) Unternehmensperformance erwartet. Sowohl für H1a ($\beta = 0,406$; p = 0,019; 90 % KI = [0,119; 0,674]) als auch für H1c ($\beta = 0,310$; p = 0,048; 90 % KI = [0,040; 0,551]) wird ein positiver und signifikanter Einfluss festgestellt, weshalb beide Hypothesen angenommen werden können. Für H1b ($\beta = -0,323$; p = 0,033; 90 % KI = [-0,559; -0,064]) wird entgegen der Erwartungen ein signifikant negativer Effekt beobachtet, wodurch diese Hypothese als verworfen gilt. Mit dem zweiten Hypothesenset (H2a, H2b und H2c) wurde ein positiver Zusammenhang zwischen NPP und den drei Dimensionen der Unternehmensperformance postuliert. Es findet sich ein positiver und signifikanter Zusammenhang zwischen NPP und der ökonomischen Unternehmensperformance, wonach H2b ($\beta = 0,211$; p = 0,052; 90 % KI = [0,040; 0,395]) angenommen werden kann. H2a ($\beta = 0,154$; p = 0,304; 90 % KI = [-0,108; 0,385]) sowie H2c ($\beta = 0,114$; p = 0,349; 90 % KI = [-0,110; 0,292]) haben zwar einen leicht positiven Einfluss, welcher aber nicht empirisch gestützt werden konnte, weshalb beide Hypothesen verworfen werden. Mit dem dritten Hypothesenset (H3a, H3b und H3c) wurde erwartet, dass IR die Unternehmensperformance positiv begünstigt. Jedoch kann für alle drei Hypothesen kein signifikanter Einfluss festgestellt

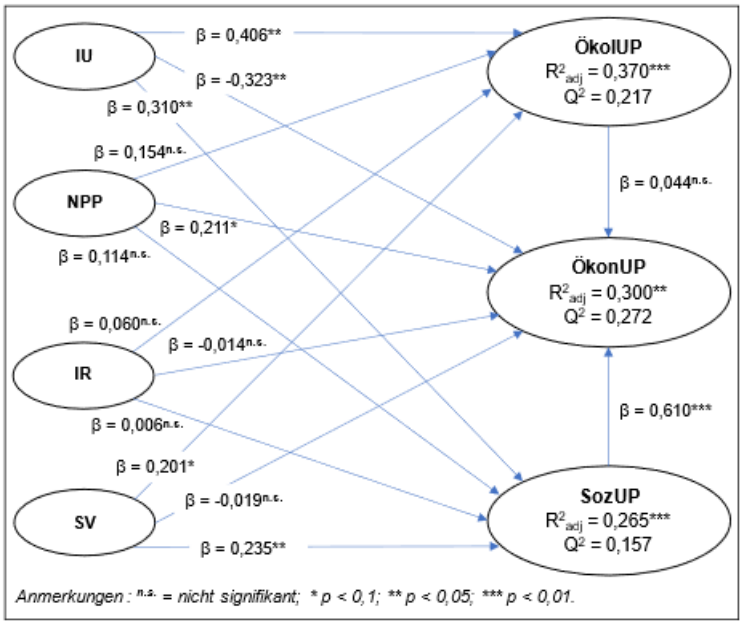


Abbildung 5: Ergebnisse des Strukturmodells von SM1

Quelle: eigene Darstellung, berechnet mit SmartPLS 3.3.2 (Ringle et al., 2012)

werden. Für SV wurde ebenfalls ein positiver Zusammenhang mit der Unternehmensperformance erwartet. SV hat sowohl auf die ökologische als auch auf die soziale Unternehmensperformance einen signifikant positiven Einfluss. Demnach werden sowohl H4a (β = 0,201; p = 0,093; 90 % KI = [-0,018; 0,368])[10] als auch H4c (β = 0,235; p = 0,038; 90 % KI = [0,034; 0,406]) angenommen. Ein positiver Zusammenhang zwischen SV und der ökonomischen Unternehmensperformance konnte nicht bestätigt werden, weshalb H4b (β = -0,019; p = 0,884; 90 % KI = [-0,221; 0,193]) verworfen wird. Die Hypothesen H9a und H9b, welche sowohl in SM1 als auch in SM2 vorkommen, postulieren einen direkten und positiven Einfluss der ökologischen und sozialen Dimension auf die ökonomische Unternehmensperformance. Während die ökologische Unternehmensperformance keinen signifikanten Einfluss hat, hat die soziale Unternehmensperformance einen stark positiven und signifikanten Einfluss auf die ökonomische Unternehmensperformance. Anhand dieser Feststellungen wird H9a (β = 0,044; p = 0,803; 90 % KI = [-0,264; 0,319]) verworfen und H9b (β = 0,610; p = 0,000; 90 % KI = [-0,264; 0,319]) angenommen.

In SM2 wurden Hypothesen zum Zusammenhang externer SSCM Praktiken zur Unternehmensperformance im Sinne der TBL aufgestellt. Für die SSCM Praktik GD wurde ein positiver Zusammenhang mit allen drei Dimensionen erwartet. GD hat einen positiven und signifikanten Einfluss auf die soziale Dimension, sodass H5c (β = 0,335; p = 0,051; 90 % KI = [0,075; 0,637]) angenommen werden kann. Zur ökologischen und ökonomischen Dimension kann kein signifikant positiver Effekt festgestellt werden, sodass H5a (β = 0,112; p = 0,556; 90 % KI = [-0,212; 0,412]) und H5b (β = -0,084; p = 0,668; 90 % KI = [-0,389; 0,251]) verworfen werden. LAÜ hat einen positiven und signifikanten Effekt auf die ökologische Unternehmensperformance. Somit kann H6a (β = 0,338; p = 0,048; 90 % KI = [0,017; 0,584]) bestätigt werden. Entgegen der Erwartungen hat LAÜ einen negativen Einfluss auf die ökonomische Unternehmensperformance, welcher aber nicht signifikant ist, weshalb H6b (β = -0,295; p = 0,144; 90 % KI = [-0,637; 0,024]) verworfen wird. Auch H6c (β = 0,272; p = 0,199; 90 % KI = [-0,102; 0,576]) muss abgelehnt werden, da der positive Zusammengang zwischen LAÜ und der sozialen Dimension nicht statistisch signifikant ist. Für ZL wurde ein positiver Effekt auf die Unternehmensperformance postuliert. Es konnte ein signifikant positiver Effekt auf die ökonomische Dimension festgestellt werden, sodass H7b (â = 0,258; p = 0,080; 90 % KI = [0,016; 0,497]) angenommen wird. Weder der leicht positive Effekt von ZL auf die ökologische noch auf die soziale Dimension ist signifikant. Demnach werden H7a (β = 0,151; p = 0,354; 90 % KI = [-0,099; 0,433]) und H7c (β = 0,147; p = 0,435; 90 % KI = [-0,158; 0,454]) verworfen. Für den Zusammenhang zwischen der SSCM Praktik NB und den drei Dimensionen der Unternehmensperformance können keine signifikant positiven Effekte festgestellt werden. Folglich werden H8a (β = 0,056 ; p = 0,711; 90 % KI = [-0,195; 0,300]), H8b (β = 0,107; p = 0,525; 90 % KI = [-0,163; 0,381]) sowie H8c (β = -0,118 ; p = 0,579; 90 % KI = [-0,435; 0,211]) nicht bestätigt. Analog zu SM1 sind die Ergebnisse für den Einfluss der ökologischen und sozialen Dimension auf die ökonomische Unternehmensperformance ähnlich. Während H9a (β = 0,009; p = 0,962; 90 % KI = [-0,303; 0,325]) keinen

[10] Obwohl das 90 % Bca-Konfidenzintervall knapp den Wert 0 enthält, wird H4a aufgrund p < 0,10 angenommen.

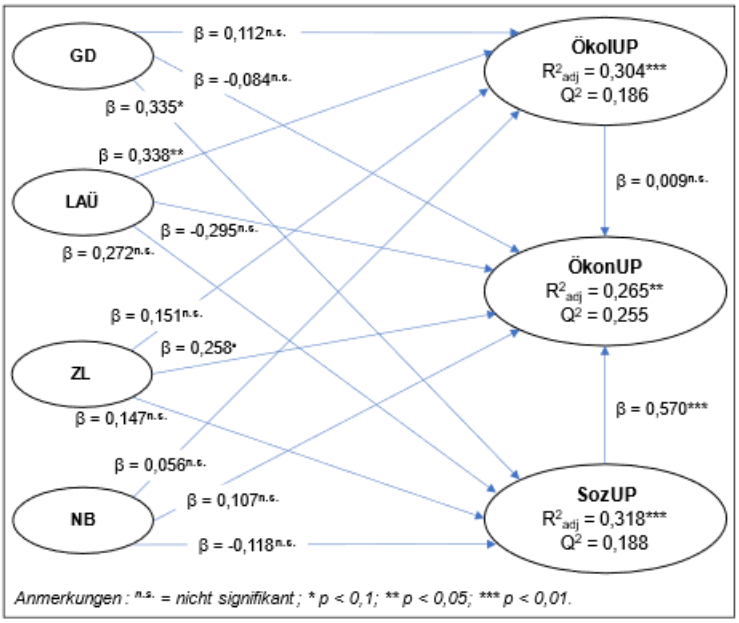


Abbildung 6: Ergebnisse des Strukturmodells von SM2

Quelle: eigene Darstellung, berechnet mit SmartPLS 3.3.2 (Ringle et al., 2012)

signifikant positiven Einfluss aufweist und deshalb verworfen wird, wird mit H9b (β = 0,570; p = 0,001; 90 % KI = [0,285; 0,864]) ein stark positiver und signifikanter Einfluss der sozialen auf die ökonomische Dimension festgestellt. In Tabelle 10 werden die Ergebnisse aller Hypothesen zusammengefasst. In SM1 werden 6 von 14 und in SM2 4 von 14 Hypothesen angenommen.

Prüfung mediierender Effekte

Da die endogenen latenten Konstrukte ÖkolUP und SozUP auch als Mediatorvariable zwischen den exogenen latenten Konstrukten (interne und externe SSCM Praktiken) und dem endogenen latenten Konstrukt ÖkonUP fungieren, müssen Mediatoreffekte bei der Datenanalyse berücksichtigt werden. In Tabelle M.1 im Anhang M werden die verschiedenen Arten von Mediatoreffekten erläutert. Ändert sich ein exogenes Konstrukt, so kann dies zur Änderung der Mediatorvariable und schlussendlich zur Änderung des endogenen Konstrukts führen (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 195–201). Die Mediatoranalyse wird gemäß der Vorgehensweise, wie sie in Abbildung M.1 im Anhang M dargestellt wird, und nachfolgend zur Veranschaulichung am Beispiel der Beziehung von IU zu ÖkonUP durchgeführt. Zuerst wird überprüft, ob die indirekten Effekte von IU auf ÖkonUP über die Mediatorvariablen ÖkolUP und SozUP im 90 % Bca-Konfidenzintervall signifikant sind. Der indirekte Effekt errechnet sich aus dem Produkt der standardisierten Pfadkoeffizienten (z.B. für IU -> SozUP -> ÖkonUP: 0,310 x 0,610 = 0,189). In Tabelle M.2 (für SM1) und Tabelle M.3 (für SM2) im Anhang M werden die Signifikanzen der direkten und indirekten Effekte dargestellt. Während der indirekte Effekt über die Mediatorvariable ÖkolUP nicht signifikant ist (β = 0,018; 90 % KI = [-0,086; 0,139]), weist der indirekte Effekt über die Mediatorvariable SozUP Signifikanz auf (β = 0,189; 90 % KI = [0,026; 0,445]). Im zweiten Schritt wird überprüft, ob der direkte Effekt von IU auf ÖkonUP signifikant ist. Dem kann zugestimmt werden (β = -0,323; 90 % KI = [-0,559; -0,064]). Da der direkte Effekt signifikant, der indirekte Effekt von IU über ÖkolUP auf ÖkonUP aber nicht signifikant ist, wird geschlussfolgert, dass hier eine „Nicht-Mediation nur mit direktem Effekt“ vorliegt (vgl. Tabelle M.4 im Anhang M). Da der indirekte Effekt über die Mediatorvariable SozUP Signifikanz aufweist, wird im letzten Schritt überprüft, ob das Produkt aller standardisierten Pfadkoeffizienten des direkten und indirekten Effekts positiv oder negativ ist. In diesem Fall ist das Produkt negativ (-0,321 x 0,310 x 0,610 = -0,061), was auf eine „kompetitive Mediation“ der Mediatorvariable SozUP hinweist. Gemäß dem Vorgehen in Abbildung M.1 im Anhang M wird also in SM1 für die Beziehung IU und ÖkonUP eine „kompetitive Mediation“ über die Mediatorvariable SozUP sowie für die Beziehung SV und ÖkonUP über SozUP eine „ausschließlich indirekte Mediation“ identifiziert. In SM2 besteht eine „ausschließlich indirekte Mediation“ für die Beziehung von SV und ÖkonUP ebenfalls über die Mediatorvariable SozUP. Eine Zusammenfassung der Ergebnisse der Mediatoranalyse wird in Tabelle M.4 im Anhang M dargestellt.

Prüfung des Bestimmtheitsmaßes und der f^2-Effektstärken

Um die Erklärungskraft des PLS-Strukturmodells zu beurteilen, wird für jedes latente endogene Konstrukt das Bestimmtheitsmaß (R^2-Wert) bestimmt. Der R^2-Wert, welcher

Tabelle 10: Hypothesenergebnisse

Submodell 1 (interne Praktiken)		Submodell 2 (externe Praktiken)	
Hypothese	**Ergebnis**	**Hypothese**	**Ergebnis**
H1a: IU -> ÖkolUP	angenommen	**H5a:** GD -> ÖkolUP	abgelehnt
H1b: IU -> ÖkonUP	abgelehnt	**H5b:** GD -> ÖkonUP	abgelehnt
H1c: IU -> SozUP	angenommen	**H5c:** GD -> SozUP	angenommen
H2a: NPP -> ÖkolUP	abgelehnt	**H6a:** LAÜ -> ÖkolUP	angenommen
H2b: NPP -> ÖkonUP	angenommen	**H6b:** LAÜ -> ÖkonUP	abgelehnt
H2c: NPP -> SozUP	abgelehnt	**H6c:** LAÜ -> SozUP	abgelehnt
H3a: IR -> ÖkolUP	abgelehnt	**H7a:** ZL -> ÖkolUP	abgelehnt
H3b: IR -> ÖkonUP	abgelehnt	**H7b:** ZL -> ÖkonUP	angenommen
H3c: IR -> SozUP	abgelehnt	**H7c:** ZL -> SozUP	abgelehnt
H4a: SV -> ÖkolUP	angenommen	**H8a:** NB -> ÖkolUP	abgelehnt
H4b: SV -> ÖkonUP	abgelehnt	**H8b:** NB -> ÖkonUP	abgelehnt
H4c: SV -> SozUP	angenommen	**H8c:** NB -> SozUP	abgelehnt
H9a: ÖkolUP -> ÖkonUP	abgelehnt	**H9a:** ÖkolUP -> ÖkonUP	abgelehnt
H9b: SozUP -> ÖkonUP	angenommen	**H9b:** SozUP -> ÖkonUP	angenommen

Quelle: eigene Darstellung

auf einen Wertebereich zwischen 0 und 1 standardisiert ist (Krafft, Götz & Liehr-Gobbers 2005, S. 83), gibt an, „wie viel Prozent der Varianz einer latent endogenen Variablen über die ihr zugeordneten unabhängigen (exogenen) Variablen erklärt wird“ (Weiber & Mühlhaus, 2010, S. 256). Da der R^2-Wert von der Anzahl an latenten exogenen Variablen im Strukturmodell beeinflusst wird, sollte aus Vergleichbarkeitsgründen das adjustierte Bestimmtheitsmaß (R^2_{adj}) zur Beurteilung der Erklärungskraft herangezogen werden (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 171). Nach Chin (1998b, S. 323) gelten Werte für R^2_{adj} ab 0,67 als substantiell, ab 0,33 als moderat und ab 0,19 als schwach. Gemäß dieser Einordnung wird nach Tabelle N.1 im Anhang N festgestellt, dass in SM1 die Erklärungskraft des latenten endogenen Konstrukts ÖkolUP (R^2_{adj} = 0,370; p = 0,000; 90 % KI = [0,153; 0,471]) als moderat und von ÖkonUP (R^2_{adj} = 0,300; p = 0,013; 90 % KI = [0,069; 0,423]) sowie SozUP (R^2_{adj} = 0,265; p = 0,009; 90 % KI = [0,055; 0,378]) als schwach angesehen werden kann. In SM2 weisen alle latenten endogenen Konstrukte ÖkolUP (R^2_{adj} = 0,304; p = 0,002; 90 % KI = [0,100; 0,414]), ÖkonUP (R^2_{adj} = 0,265; p = 0,011; 90 % KI = [0,060; 0,351]) und SozUP (R^2_{adj} = 0,318; p = 0,004; 90 % KI = [0,111; 0,456]) eine schwache Erklärungskraft auf (vgl. Tabelle N.2 im Anhang N).

Zusätzlich zur Prüfung der R^2_{adj}-Werte der latenten endogenen Konstrukte kann anhand der f^2-Effektstärken überprüft werden, welchen Effekt einzelne latente exogene Variablen auf die Erklärungskraft der latenten endogenen Konstrukte haben. Dazu wird die Veränderung des R^2-Wertes für ein bestimmtes endogenes Konstrukt evaluiert, wenn das betrachtete exogene Konstrukt bei der Schätzung des PLS-Strukturmodells nicht berücksichtigt wird (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 173). Nach Chin (1998b, S. 316) sowie Weiber und Mühlhaus (2010, S. 257) errechnet sich die f^2-Effektstärke wie folgt:

$$f^2_{ij} = \frac{R^2_{inkl.} - R^2_{exkl.}}{1 - R^2_{inkl.}} \quad (1)$$

mit $R^2_{inkl.}$ = R^2-Wert der endogenen Variablen j, sofern alle exogenen Variablen zur Schätzung herangezogen werden.

$R^2_{exkl.}$ = R^2-Wert der endogenen Variablen j, sofern die exogene Variable i nicht zur Schätzung herangezogen wird.

Je deutlicher sich der R^2-Wert durch diese Prozedur verschlechtert, desto höher ist der Effekt des betrachteten exogenen Konstrukts (Weiber & Mühlhaus, 2010, S. 257). Zur Beurteilung der f^2-Effektstärken werden die von Cohen (1988, S. 413–414) vorgeschlagenen Richtwerte herangezogen. Demnach repräsentieren Werte ab 0,02 einen kleinen, ab 0,15 einen mittleren und ab 0,35 einen großen Effekt des latenten exogenen Konstrukts. f^2-Effektstärken mit Werten kleiner als 0,02 haben somit keinen Einfluss (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 173). Die mithilfe des PLS-Algorithmus von SmartPLS 3.3.2 errechneten f^2-Effektstärken können Tabelle N.3 im Anhang N entnommen werden. In SM1 hat z.B. IU einen kleinen Effekt auf ÖkolUP (f^2 = 0,142), ÖkonUP (f^2 = 0,073) und SozUP (f^2 = 0,071). Ein mittelgroßer Effekt wirkt sich von SozUP auf ÖkonUP (f^2 = 0,313) aus. In SM2 hat SozUP ebenfalls einen mittelgroßen Effekt auf ÖkonUP (f^2 = 0,246).

Prüfung der Prognoserelevanz und der q^2-Effektstärken

Weiterhin gilt es die Prognoserelevanz der latenten endogenen Konstrukte (Q^2-Wert bzw. Stone-Geisser-Kriterium) der PLS-Strukturmodelle zu überprüfen (Weiber & Mühl-

haus, 2010, S. 257). Der Q^2-Wert wird mithilfe der sogenannten Blindfolding-Prozedur ermittelt, bei welcher „während der Parameterschätzung systematisch ein Teil der Urdatenmatrix als fehlend angenommen [...] und anschließend mit den berechneten Parameterwerten die als fehlend angenommenen Rohdaten wieder prognostiziert“ wird (Weiber & Mühlhaus, 2010, S. 258). Dieser iterative Prozess wird so lange durchgeführt, bis jeder Datenpunkt der Urdatenmatrix einmal ausgelassen wurde (Hair, Hult, Ringle & Sarstedt, 2017, S. 202). Vor Durchführung der Blindfolding-Prozedur muss die sogenannte Auslassungsdistanz (D) bestimmt werden. Die Auslassungsdistanz gibt an, welcher „D-te“ Datenpunkt der Urdatenmatrix während des Blindfoldings ausgelassen werden soll. Dabei sollte die Anzahl der Beobachtungen (hier N = 61) dividiert durch die Auslassungsdistanz keine ganze Zahl ergeben, da ansonsten in jeder Iteration das gleiche Set an Datenpunkten aus der Urdatenmatrix ausgelassen werden würde. Demnach wird für die vorliegenden Strukturmodelle eine Auslassungsdistanz von D = 7 gewählt (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 176). Ein Q^2-Wert größer 0 impliziert, dass das Modell eine hinreichende Prognoserelevanz ausweist (Chin, 1998b, S. 318). Die Ergebnisse der Blindfolding-Prozedur, die mithilfe von SmartPLS 3.3.2 durgeführt wurde, können Tabelle O.1 im Anhang O entnommen werden. Für alle latenten endogenen Konstrukte von SM1 ($Q^2_{\text{ÖkolUP}}$= 0,217; $Q^2_{\text{ÖkonUP}}$ = 0,272; Q^2_{SozUP} = 0,157) und SM2 ($Q^2_{\text{ÖkolUP}}$ = 0,186; $Q^2_{\text{ÖkonUP}}$ = 0,255; Q^2_{SozUP} = 0,188) konnten Q^2-Werte von größer 0 ermittelt werden, was auf eine hinreichende Prognoserelevanz beider Modelle hinweist.

Analog zu den f^2-Effektstärken für das Bestimmtheitsmaß kann auch bei der Prognoserelevanz mithilfe der q^2-Effektstärken ermittelt werden, welche latenten exogenen Konstrukte die stärksten Effekte auf die Prognoserelevanz der latenten endogenen Konstrukte haben (Hair, Hult, Ringle & Sarstedt, 2017, S. 207). Dazu wird die Veränderung des Q^2-Wertes für ein bestimmtes endogenes Konstrukt evaluiert, wenn das betrachtete exogene Konstrukt bei der Durchführung der Blindfolding-Prozedur nicht berücksichtigt wird (Weiber & Mühlhaus, 2010, S. 258). Nach Chin (1998b, S. 318) sowie Weiber und Mühlhaus (2010, S. 258) gilt folgende Gleichung zur Berechnung der q^2-Effektstärken:

$$q^2_{ij} = \frac{Q^2_{inkl.} - Q^2_{exkl.}}{1 - Q^2_{inkl.}} \quad (2)$$

mit $Q^2_{inkl.}$ = Q^2-Wert der endogenen Variablen j, sofern alle exogenen Variablen zur Prognose der fehlenden Urdaten herangezogen werden.

$Q^2_{exkl.}$= Q^2-Wert der endogenen Variablen j, sofern die exogene Variable i nicht zur Prognose der fehlenden Urdaten herangezogen wird.

Wie bei der Evaluierung der f^2-Effektstärken gelten auch hier dieselben Richtwerte mit 0,02 (schwacher Effekt), 0,15 (mittlerer Effekt) und 0,35 (starker Effekt) (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017, S. 178). Da SmartPLS 3.3.2 die Berechnung der q^2-Effektstärken nicht unterstützt, wurden diese manuell berechnet. Dazu wurde für SM1 bzw. SM2 nacheinander jedes exogene latente Konstrukt einmal ausgelassen, die Blindfolding-Prozedur durchgeführt und gemäß Gleichung 2 die entsprechende q^2-Effektstärke ermittelt. Die Ergebnisse können Tabelle O.1 im Anhang O entnommen werden. Es konnte ein mittlerer Effekt von SozUP auf die Prognoserelevanz ÖkonUP für SM1 (q^2 = 0,235) und SM2 (q^2 = 0,185) festgestellt werden. Für die anderen exogenen latenten Konstrukte wurden keine bis schwache q^2-Effektstärken ermittelt. Die negativen q^2-Effektstärken sind dabei nicht überraschend, da die zugehörigen standardisierten Pfadkoeffizienten keinen signifikanten Einfluss zeigten (vgl. Prüfung der Pfadkoeffizienten der Strukturpfade in diesem Kapitel) (Matzler et al., 2015, S. 127).

Prüfung der Kontrollvariable der Unternehmensgröße

Wie in der Unternehmensforschung üblich, wurde weiterhin überprüft, ob die Unternehmensgröße gemessen am Umsatz des abgelaufenen Geschäftsjahres einen signifikanten Einfluss auf die endogenen latenten Konstrukte ÖkolUP, ÖkonUP sowie SozUP hat. Dazu wurde der Jahresumsatz als weiteres latentes exogenes Konstrukt in beide Submodelle eingefügt und mit den endogenen latenten Konstrukten verbunden. Der Jahresumsatz wurde anhand eines Single-Items auf einer siebenstufigen Skala gemessen (vgl. Tabelle A.1 im Anhang A). Um zu testen, ob ein signifikanter Einfluss der Kontrollvariable vorliegt, wurde erneut der PLS-Algorithmus sowie das Bootstrapping-Verfahren mittels SmartPLS 3.3.2 (Ringle et al., 2020) angewendet. Es konnte weder in SM1 noch in SM2 ein signifikanter Einfluss der Unternehmensgröße auf die endogenen latenten Konstrukte festgestellt werden (vgl. Tabelle 11).

Prüfung auf Unterschiede im Implementierungsgrad interner und externer Praktiken

Weiterhin wurde überprüft, ob signifikante Unterschiede beim Grad der Implementierung von internen und externen SSCM Praktiken vorliegen. Dazu wurden die Mittelwerte aller 20 Indikatoren der internen mit allen 16 Indikatoren der externen SSCM Praktiken (N = 36) anhand eines zweiseitigen t-Tests auf Mittelwertgleichheit für unabhängige Stichproben verglichen. Da die Indikatoren auf einer Fünf-Punkte-Likert-Skala mit *(1) stimme überhaupt nicht zu* bis *(5) stimme voll zu* gemessen wurden, impliziert ein höherer Mittelwert des Indikators einen stärkeren Grad der Implementierung der jeweiligen SSCM Maßnahme. Wie Tabelle P.1 im Anhang P entnommen werden kann, ist der Grad der Implementierung interner SSCM Praktiken (M = 4,110; SD = 0,553) in den befragten Unternehmen höher, als bei externen SSCM Praktiken (M = 3,244; SD = 0,602). Die Voraussetzung von Varianzhomogenität beider Gruppen für den t-Test konnte anhand des Levene-Tests der Varianzgleichheit bestätigt werden (F(1,34) = 0,100; p = 0,754; N = 53). Der t-Test zeigt, dass es einen statistisch signifikanten Unterschied zwischen dem Grad der

Tabelle 11: Prüfung der Kontrollvariable

Strukturpfade	β	SD	T-Werte	P-Werte	90 % Bca-Konfidenzintervall
Submodell 1 (interne SSCM Praktiken)					
Jahresumsatz -> ÖkolUP	-0,139$^{n.s.}$	0,110	1,266	0,206	[-0,316; 0,044]
Jahresumsatz -> ÖkonUP	0,031$^{n.s.}$	0,111	0,281	0,779	[-0,136; 0,230]
Jahresumsatz -> SozUP	-0,139$^{n.s.}$	0,120	1,154	0,249	[-0,332; 0,064]
Submodell 2 (externe SSCM Praktiken)					
Jahresumsatz -> ÖkolUP	-0,160$^{n.s.}$	0,110	1,449	0,147	[-0,335; 0,032]
Jahresumsatz -> ÖkonUP	0,107$^{n.s.}$	0,112	0,948	0,343	[-0,090; 0,279]
Jahresumsatz -> SozUP	-0,158$^{n.s.}$	0,132	1,195	0,232	[-0,374; 0,057]

Anmerkungen: $\boldsymbol{\beta}$ *= standardisierte Pfadkoeffizienten;* ***SD*** *= Standardabweichung;* $^{n.s.}$ *= nicht signifikant.*
Quelle: eigene Darstellung, berechnet mit SmartPLS 3.3.2 (Ringle et al., 2020)

Implementierung von internen und externen SSCM Praktiken gibt. Entsprechend Tabelle P.2 im Anhang P ist der Implementierungsgrad bei internen SSCM Praktiken durchschnittlich ca. 0,866 Punkte höher, als bei externen SSCM Praktiken ($t(34) = 4{,}492$; $p = 0{,}000$; 90 % KI = [0,540; 1,192]).

5. Diskussion der Ergebnisse

Die im vorherigen Kapitel durchgeführte Datenanalyse und Präsentation der Ergebnisse bildet die Grundlage für die kritische Auseinandersetzung mit den Ergebnissen, die Implikationen für die Praxis sowie für Limitationen der vorliegenden Arbeit und das daraus resultierende Forschungspotenzial. Im ersten und zweiten Teilkapitel werden die Ergebnisse der PLS-SGM für SM1 und SM2 evaluiert und interpretiert. Darauf aufbauend werden im dritten Abschnitt dieses Kapitels Implikationen für die Praxis gegeben. Abschließend werden Limitationen aufgezeigt und es wird ein Ausblick für weiteres Forschungpotenzial gegeben.

5.1. Interne SSCM Praktiken und die Unternehmensperformance

Internes Umweltmanagement

Bei der Analyse der standardisierten Pfadkoeffizienten konnte für die interne SSCM Praktik IU wie erwartet ein positiver und statistisch signifikanter Zusammenhang mit der ökologischen Unternehmensperformance festgestellt werden ($\beta = 406$), wonach die Hypothese H1a bestätigt werden konnte. Diese Erkenntnis geht z. B. auch mit den Ergebnissen von Das (2018, S. 5788), J. Wang und Dai (2018, S. 10) oder Zhu und Sarkis (2004, S. 282) einher. Es ist nicht überraschend, dass IU einen positiven Zusammenhang mit der ökologischen Unternehmensperformance aufweist. So führt die Implementierung von UMS, wie z. B. die ISO-Normreihe 14000 dazu, dass der Ausstoß von THG oder anderen giftigen Stoffen reduziert wird oder Ressourcen effizienter eingesetzt werden (ISO, 2015, S. 2). Umweltschulungen und eine strikte Prüfung der Einhaltung von Umweltschutznormen und -richtlinien von Seiten des Unternehmens sensibilisieren zudem die Mitarbeiter*innen, sich gemäß den Vorschriften zu verhalten und somit zur Verbesserung der Öko-Effektivität des Unternehmens beizutragen (Florida, 1996, S. 92–93).

Entgegen der Erwartungen und der Ergebnisse von z. B. J. Wang und Dai (2018, S. 10) oder Zhu und Sarkis (2004, S. 277) konnte kein positiver Einfluss von IU auf die ökonomische Unternehmensperformance festgestellt werden, weshalb H1b abgelehnt werden musste. Stattdessen wurde ein signifikant negativer Effekt nachgewiesen (β = -0,323). Dies könnte daran liegen, dass Investitionen in die Implementierung von UMS sehr kapitalaufwändig sind und sich dementsprechend vorerst negativ auf die ökonomische Unternehmensperformance auswirken, bevor sich die Investitionen rentieren (J. Wang & Dai, 2018, S. 13–14). Da die Befragung eine Momentaufnahme der befragten Unternehmen darstellt, könnte eine zukünftige Befragung nach der Etablierung von UMS zu den erwarteten Ergebnissen führen.

Für den Einfluss von IU auf die soziale Unternehmensperformance konnte gemäß den Erwartungen und den Ergebnissen von J. Wang und Dai (2018, S. 10) sowie Yildiz Çankaya und Sezen (2019, S. 113) ein positiver (β = 0,310) und signifikanter Einfluss festgestellt werden, weshalb H1c angenommen wurde. So lässt sich die Vermutung bestätigen, dass die Einführung von internen UMS die Gesundheit und die Sicherheit der Mitarbeiter*innen am Arbeitsplatz verbessert und zur Reduzierung von Krankheitstagen führt, da die Mitarbeiter*innen z. B. weniger mit giftigen Stoffen in Berührung kommen oder weniger Emissionen ausgesetzt sind. In Bezug auf die Institutional Theory und die Stakeholder Theory (vgl. Kapitel 2.3.2) wirkt sich dies zudem positiv auf die Unternehmensreputation aus Sicht aller internen und externen Stakeholder aus.

Während kein positiv-direkter Einfluss von IU auf die ökonomische Unternehmensperformance festgestellt wurde, konnte bei der Mediatoranalyse ein signifikanter und positiv-indirekter Effekt (β = 0,189) von IU über SozUP auf ÖkonUP entdeckt werden. Wie zuvor erwähnt, verbessert IU die soziale Unternehmensperformance. Dies kann entsprechend der RBT nach Barney (1991, S. 99–101) und des

NRBV nach Hart (1995, S. 991) aufgrund der verbesserten Unternehmensreputation und des höheren Produktimages zu wertvollen Wettbewerbsvorteilen führen, was sich wiederum positiv auf die ökonomische Unternehmensperformance, wie z. B. durch eine Umsatzsteigerung auswirken kann.

Nachhaltiges Produkt- und Prozessdesign

Das Ziel der zweiten internen SSCM Praktik NPP ist es, die Produkte und Produktionsprozesse der Unternehmen nachhaltiger zu gestalten. Es konnte ein leicht-positiver, aber nicht signifikanter Zusammenhang zwischen NPP und der ökologischen Unternehmensperformance festgestellt werden (β = 0,154), wonach Hypothese H2a entgegen der Erwartungen und den Ergebnissen von z. B. Paulraj et al. (2017, S. 253) verworfen werden musste. Obwohl der leicht-positive Zusammenhang nicht statistisch belegbar ist, kann trotzdem grundsätzlich davon ausgegangen werden, dass NPP sich positiv auf die ökologische Unternehmensperformance auswirkt, da sich durch nachhaltige Produktionsprozesse z. B. der Verbrauch an Ressourcen verringert. Dieser Zusammenhang kann auch durch die niedrige f^2-Effekstärke (f^2 = 0,031) erklärt werden. Möglicherweise ist der leicht-positive Einfluss aufgrund des Rebound-Effekts nicht signifikant. Die mit der Implementierung von NPP verbundene Steigerung der Öko-Effizienz könnte im schlimmsten Fall dazu führen, dass durch die effizienten Produktionsprozesse und den damit einhergehenden niedrigeren Herstellungskosten der Produkte das Konsumverhalten angeregt wird und sich damit der positive Einfluss auf die ökologische Unternehmensperformance relativiert (vgl. Kapitel 2.1.2).

Weiterhin wurde erwartungsgemäß ein signifikanter und positiver Einfluss von NPP auf die ökonomische Dimension festgestellt (β = 0,211), was dem Ergebnis von Paulraj et al. (2017, S. 253) entspricht. Die zugehörige Hypothese H2b konnte deshalb angenommen werden. Ein effizientes Prozessdesign führt zur Reduzierung von Abfällen und Emissionen. Weiterhin erlaubt der Einsatz neuartiger Technologien Einsparungen von Energie und Wasser, was wiederum zu Kosteneinsparungen und dementsprechend zu einer Verbesserung der ökonomischen Unternehmensperformance führen kann (Rao & Holt, 2005, S. 902–903; Hsu et al., 2013, S. 673). In Bezug auf RBT und NRBV können nachhaltige Innovationen im Produkt- und Prozessdesign nicht imitierbare und nicht substituierbare Ressourcen und Fähigkeiten darstellen, die wiederum zu Wettbewerbsvorteilen führen und somit die ökonomische Unternehmensperformance positiv beeinflussen können (Kleindorfer et al., 2005, S. 485).

Es wurde ein leicht-positiver Einfluss von NPP auf die soziale Unternehmensperformance gefunden, welcher sich aber als statistisch nicht signifikant herausstellte (β = 0,114), weshalb H2c verworfen wurde. Grundsätzlich kann argumentiert werden, dass sich NPP durch saubere und sichere Produktionsprozesse positiv auf die Unternehmensreputation und das Produktimage sowie auf das Wohlbefinden der Mitarbeiter*innen und der umliegenden Gemeinden auswirken (J. Wang & Dai, 2018, S. 5; Yildiz Çankaya & Sezen, 2019, S. 113), was aber nicht anhand des theoretischen Forschungsmodells und der empirischen Daten so bestätigt werden konnte. Dies liegt womöglich daran, dass das latente exogene Konstrukt NPP anhand seiner Indikatoren mehr eine ökologische und ökonomische, als eine soziale SSCM Praktik darstellt.

Investment Recovery

Entgegen der Erwartungen konnte für IR kein signifikanter Einfluss auf die ökologische (β = 0,060) Dimension festgestellt werden, weshalb H3a abgelehnt werden musste. Dieses Ergebnis deckt sich mit keiner der untersuchten Studien von z. B. Zhu und Sarkis (2004, S. 281), Esfahbodi et al. (2017, S. 24) oder Y. Lu et al. (2018, S. 20). Die deskriptiven Statistiken (vgl. Tabelle B.1 im Anhang B) der Umfrageergebnisse zeigen, dass für alle drei Indikatoren von IR der Mittelwert geringer im Vergleich zu den anderen internen SSCM Praktiken ist (M_{IR_1} = 3,410; M_{IR_2} = 2,790; M_{IR_3} = 3,520), wobei besonders der Indikator IR_2 vergleichsweise auffällig niedrig von den Befragten bewertet wird. IR_2 beinhaltet den Verkauf defekter und gebrauchter Materialien. Der größte Anteil der Befragten stammt aus den Bereichen Metallerzeugung und -bearbeitung, Herstellung von Metallerzeugnissen, Herstellung von Nahrungs- und Genussmitteln, Getränken und Tabakerzeugnissen sowie Maschinenbau (vgl. Tabelle 4). Für den Bereich Metallerzeugung sowie Maschinenbau kann vermutet werden, dass defekte und gebrauchte Materialien eher recycelt und wiederverwertet werden, anstatt verkauft zu werden. Aber auch beim Recycling von Materialien können Schwierigkeiten auftreten, da z. B. der Gebrauch von verschiedenen Verbundstoffen in Altmetalllegierungen das Wiederverwertungspotential einschränkt oder unmöglich macht. Auch die Verwertung elektrischer Geräte und Maschinen stellt sich schwierig dar, weil diese oftmals aus vielen verschiedenen Materialien bestehen, was die Sortierung dieser aufwändig gestaltet (OECD, 2006[11], zitiert nach Wilts, Lucas, von Gries & Zirngiebl, 2014, S. 45). In der Lebensmittelbranche ist aufgrund hoher Regularien und Hygienestandards die Wiederverwendbarkeit z. B. von verdorbenen Lebensmitteln nicht möglich, weshalb diese entsorgt werden müssen. Auch die Unternehmen der Stichprobe könnten ein Grund dafür sein, dass IR keinen signifikanten Einfluss auf die ökologische Unternehmensperformance hat, wenn zufällig größtenteils Unternehmen mit einer niedrigen Recyclingquote an der Befragung teilgenommen haben.

Auch auf die ökonomische Unternehmensperformance hat IR keinen signifikanten Einfluss (β = -0,014), weshalb H3b verworfen wurde. Dies könnte dem Fakt geschuldet sein, dass die Investitionen z. B. in Recyclinganlagen und -technologien sehr kapitalaufwändig sind und diese sich erst nach mehreren Jahren rentieren (Zhu & Sarkis, 2004, S. 283). Wie schon erwähnt, ist die Befragung als Querschnittstudie eine Momentaufnahme der Situation in den befragten

[11] OECD (2006): Improving Recycling Markets, OECD Publishing Paris, 2006.

Unternehmen. Eine weitere Befragung zu einem späteren Zeitpunkt könnte deshalb die erwarteten Ergebnisse hervorbringen.

Die soziale Unternehmensperformance wird ebenfalls nicht signifikant durch IR beeinflusst ($\beta = 0{,}006$). Dementsprechend wurde entgegen der Erwartungen und den Ergebnissen von Y. Lu et al. (2018, S. 21) sowie Yildiz Çankaya und Sezen (2019, S. 113) Hypothese H3c verworfen. Dies könnte daran liegen, dass IR eine SSCM Praktik mit Fokus auf der ökologischen und ökonomischen Dimension darstellt und die soziale Dimension eher indirekt beeinflussen könnte, indem die Unternehmensreputation durch die ökologischen und ökonomischen Aspekte des Konstrukts gesteigert wird. Ein statistischer Beweis dafür konnte durch die vorliegende Untersuchung aber nicht gefunden werden.

*Soziale Verantwortung gegenüber Mitarbeiter*innen und Gesellschaft*

Für die soziale SSCM Praktik SV konnte ein positiver und signifikanter Einfluss auf die ökologische Unternehmensperformance gezeigt werden ($\beta = 0{,}201$), wonach H4a angenommen werden konnte. Dieser Zusammenhang wurde in der gesichteten Literatur nicht überprüft und kann somit als neuartiger Zusammenhang in der SSCM-Forschung gesehen werden. Eine Erklärung für dieses Ergebnis könnte der Ausbau und die Förderung von Weiterbildungsmaßnahmen der Mitarbeiter*innen in Bezug auf umweltspezifische Aspekte sein. Die Mitarbeiter*innen werden dadurch sensibilisiert, sich gemäß den Umweltvorschriften des Unternehmens zu verhalten und somit zur Verbesserung der Öko-Effektivität beizutragen. Diese Vorschriften enthalten Regeln, wie sich die Mitarbeiter*innen in allen Funktionen zu verhalten haben, um Ressourcen, wie z. B. Wasser und Energie, einzusparen. Auch die Häufigkeit von Umweltunfällen kann durch gezieltes Training der Mitarbeiter*innen verringert werden (Florida, 1996, S. 92–93).

Die ökonomische Unternehmensperformance wird nicht signifikant und nicht direkt durch SV beeinflusst ($\beta = -0{,}019$), weshalb H4b entgegen der Erwartungen und der Ergebnisse von Zhu et al. (2016, S. 423–424) verworfen wurde. Dies könnte daran liegen, dass der Ausbau von Fortbildungsmöglichkeiten und Schulungen für die Mitarbeiter*innen sehr kostspielig ist. Weiterhin könnte die Implementierung sozialer Programme, wie z. B. flexible Arbeitszeiten bei Werksarbeitern zu Schwierigkeiten bei der Produktionsplanung und Defiziten bei der Planungssicherheit führen. Auch die Einführung von hohen Sicherheitsstandards kann z. B. dazu führen, dass einige Aktivitäten im Produktionsprozess länger dauern. Dies wirkt sich schließlich negativ auf die Stückzahl pro Stunde und dementsprechend auch auf die Herstellungskosten pro Stück aus, was einen negativen Einfluss auf die ökonomische Unternehmensperformance nach sich ziehen könnte (Gimenez et al., 2012, S. 156). Zudem besteht insbesondere für börsennotierte Unternehmen in Deutschland ab einer Größe von 500 Mitarbeiter*innen seit 2017 nach dem sogenannten „CSR-Richtlinie-Umsetzungsgesetz“ eine Pflicht zur nicht-finanziellen Berichterstattung (Bundesministerium für Arbeit und Soziales 2020). Dieser zwanghafte Druck durch staatliche Regularien (vgl. Institutional Theory in Kapitel 2.3.2) führt für die betroffenen Unternehmen zu höheren Reporting Kosten. Demnach kann angenommen werden, dass die ökonomischen Vorteile, die durch die Maßnahme SV entstehen können, wie z. B. Reduzierung von Fehlzeiten oder die Erhöhung der Arbeitsmoral durch die nötigen Investitionen, relativiert werden und somit SV keinen direkten und positiven Einfluss auf die ökonomische Unternehmensperformance hat. Da die Befragung aber eine Momentaufnahme in den befragten Unternehmen war und dementsprechend nicht die Auswirkung auf die ökonomische Unternehmensperformance auf lange Sicht berücksichtigt werden konnte, könnte eine weitere Erklärung für das Ergebnis sein, dass die Kosten für SV Praktiken auf kurze Sicht zwar hoch erscheinen, die aus SV resultierenden Vorteile aber auf lange Sicht überwiegen und zu einer Steigerung der ökonomischen Unternehmensperformance führen könnten (Cruz & Wakolbinger, 2008, S. 72).

Wie erwartet, wurde für SV ein positiver und signifikanter Effekt auf die soziale Unternehmensperformance festgestellt ($\beta = 0{,}235$). Somit konnte Hypothese H4c bestätigt werden. Dieses Ergebnis geht mit den Resultaten der Studie von Zhu et al. (2016, S. 423–424) einher. Maßnahmen, wie eine angemessene Entlohnung und Gewährung von Urlaubstagen (z. B. durch die Koppelung an einen Tarifvertrag), ein sicheres und gesundes Arbeitsumfeld, Gesundheitsleistungen oder persönliche Entwicklungschancen für alle Beschäftigten erhöhen die Zufriedenheit der Mitarbeiter*innen oder reduzieren die Fehltage und Arbeitsunfälle. Dies wirkt sich positiv auf die Sozio-Effizienz und Sozio-Effektivität und somit auf die gesamte soziale Unternehmensperformance aus. Weiterhin führen SV Praktiken, wie z. B. die regelmäßige Offenlegung sozialspezifischer Aspekte in einem CSR-Report zu einer Verbesserung der Unternehmensreputation. Auch Investitionen des Unternehmens in soziale Projekte der umliegenden Gemeinden tragen einen Teil dazu bei, dass das Image des Unternehmens verbessert wird (Zhu et al., 2016, S. 424; Gimenez et al., 2012, S. 156).

Zwar konnte kein direkter Effekt von SV auf die ökonomische Unternehmensperformance bestätigt werden, jedoch wurde bei der Mediatoranalyse ein signifikant-positiver, indirekter Effekt von SV über die soziale auf die ökonomische Unternehmensperformance festgestellt ($\beta = 0{,}143$). Neben der Erreichung ökonomischer Ziele sorgen sich heutzutage die Stakeholder auch zunehmend um sozialspezifische Aspekte (Golicic & Smith, 2013, S. 92). Gemäß der Stakeholder Theory kann eine gute Unternehmensreputation aus Sicht aller internen und externen Stakeholder dazu führen, dass auch das Produktimage steigt und somit indirekt z. B. der Umsatz oder der Marktanteil des Unternehmens durch die soziale Performance gesteigert wird (Weber, 2008, S. 249). Auch hier kann zusätzlich Bezug auf die RBT genommen werden, denn durch die erfolgreiche Umsetzung von SV werden wertvolle Wettbewerbsvorteile generiert, was wiederum einen positiven Ein-

fluss auf die ökonomische Unternehmensperformance haben kann.

5.2. Externe SSCM Praktiken und die Unternehmensperformance

Green Distribution

Es konnte ein leicht positiver, aber nicht signifikanter Einfluss von GD auf die ökologische Unternehmensperformance festgestellt werden (β = 0,112). Dementsprechend musste entgegen der Erwartungen und der Ergebnisse von Zailani et al. (2012, S. 338) die Hypothese H5a verworfen werden. Demnach scheint GD im Kontext deutscher herstellender und verarbeitender Unternehmen keine bedeutende SSCM Praktik zu sein. Gemäß der deskriptiven Statistik (vgl. Tabelle B.1 im Anhang B) lässt sich anhand der Indikatoren GD_2 (M = 3,130; Med = 3) und GD_3 (M = 3,070; Med = 3) vermuten, dass die Auswahl und die Überwachung der Transportdienstleister keine große Rolle bei den befragten Unternehmen einnimmt und sich deshalb nicht signifikant auf die ökologische Unternehmensperformance auswirkt.

Auch auf die ökonomische Unternehmensperformance konnte kein entscheidender Einfluss von GD beobachtet werden (β = -0,084), wonach H5b abgelehnt wurde. Dies kann zum einen darauf zurückzuführen sein, dass Investitionen in ökologisch nachhaltige Infrastrukturen getätigt werden müssen (Esfahbodi et al., 2017, S. 23), wie z. B. in die Entwicklung nachhaltiger Verpackungen. Zum anderen entstehen Transaktions- und Wechselkosten sowohl bei der Auswahl von neuen Logistikpartnern nach umwelt- und sozialverantwortlichen Kriterien als auch bei der Verfolgung und Überwachung der Emissionen, die bei der Distribution der Produkte durch die Logistikpartner verursacht werden. Diese Kosten könnten die finanziellen Einsparungen durch optimales Verpackungsdesign, weniger Abfallkosten oder durch die Routenoptimierung und den damit einhergehenden niedrigeren Energiekosten relativieren (Wu & Dunn, 1995, S. 29; Kafa et al., 2013, S. 72). Auch hier könnte eine weitere Befragung zu einem späteren Zeitpunkt zu dem erwarteten Ergebnis führen, wenn die Investitionen sich rentieren und sich langfristige Partnerschaften mit nachhaltigen Logistikdienstleistern etabliert haben.

Erwartungsgemäß wurde ein signifikanter und positiver Einfluss von GD auf die soziale Unternehmensperformance festgestellt (β = 0,335), wonach H5c bestätigt werden konnte. Dieses Ergebnis ist mit den Ergebnissen der Untersuchung von Zailani et al. (2012, S. 337) vereinbar. Da sich die Kundenbedürfnisse immer mehr auf nachhaltige Verpackungen konzentrieren, führt die Verwendung ökologisch nachhaltiger, recyclebarer und sicherer Verpackungen dazu, dass die Kundenzufriedenheit und somit auch die Unternehmensreputation und das Image der Produkte steigt (Zailani et al., 2012, S. 338). Auch die Auswahl der Logistikdienstleister nach umwelt- und sozialspezifischen Aspekten wirkt sich positiv auf die Reputation des Unternehmens aus. Weiterhin konnte mithilfe der Mediatoranalyse eine signifikant-positive, ausschließlich indirekte Mediation der sozialen Unternehmensperformance zwischen GD und der ökonomischen Unternehmensperformance festgestellt werden (β = 0,191). Hier kann erneut auf Basis der Stakeholder Theory argumentiert werden, dass ein verbessertes Produktimage sowie eine höhere Unternehmensreputation die Umsätze und Marktanteile der Unternehmen steigen lässt. Gemäß RBT und NRBV kann eine erfolgreich implementierte GD zu Wettbewerbsvorteilen führen und somit auch die ökonomische Unternehmensperformance positiv beeinflussen.

Lieferantenauswahl und -überwachung

Wie erwartet, wurde ein signifikanter und positiver Einfluss von LAÜ auf die ökologische Unternehmensperformance beobachtet (β = 0,338), wonach H6a bestätigt werden konnte. Wie bei der Herleitung der Hypothesen schon beschrieben wurde, zeigen nicht alle Lieferanten, besonders aus Schwellenländern, ethisches Verhalten in Bezug auf umwelt- und sozialkritische Aspekte. Durch eine strenge Auswahl und Überwachung von Lieferanten soll verhindert werden, dass unethische Zulieferer in das Lieferantenportfolio des fokalen Unternehmens gelangen (Busse, 2016, S. 442; J. Wang & Dai, 2018, S. 6). Die Ergebnisse zeigen, dass sich in den befragten Unternehmen die SSCM Praktik LAÜ etabliert und einen positiven Einfluss auf die ökologische Unternehmensperformance hat. Die fokalen Unternehmen üben gemäß der Institutional Theory (vgl. Kapitel 2.3.2) einen normativen Druck auf alle in Frage kommenden Lieferanten aus. Dazu zählen z. B. Investitionen in nachhaltige Technologien oder die Zertifizierung nach ISO-Normreihe 14000, um den hohen umweltspezifischen Anforderungen des fokalen Unternehmens gerecht zu werden und somit in deren Lieferantenportfolio aufgenommen zu werden (Klassen & Vachon, 2003, S. 347). Durch eine sorgfältige LAÜ wird verhindert, dass der Einsatz umweltbelastender Rohstoffe und Materialien in den Produktionsprozess des fokalen Unternehmens gelangen. Dies wirkt sich womöglich nicht nur positiv auf die ökologische Performance des fokalen Unternehmens aus, sondern auch auf die der Lieferanten. Nach Florida (1996, S. 93) könnte eine weitere Erklärung für den positiven Einfluss von LAÜ auf die ökologische Unternehmensperformance sein, dass fokale Unternehmen die besonders umweltbelastenden Produktionsprozesse auf ihre Lieferanten abwälzen, um die eigene Umweltbilanz zu verbessern. Diese Prozedur wäre aber nicht als besonders nachhaltig für das gesamte Supply Chain Netzwerk zu beurteilen, da die Umweltbelastung lediglich auf eine vorgelagerte Stufe verschoben wird.

Entgegen der Erwartungen wurde kein signifikanter und positiver Effekt von LAÜ auf die ökonomische Unternehmensperformance festgestellt (β = -0,295), weshalb H6b verworfen wurde. Dieses Ergebnis geht mit den Resultaten von Gimenez et al. (2012, S. 157) sowie J. Wang und Dai (2018, S. 14) einher, die auch keinen statistisch signifikanten Einfluss von LAÜ auf die ökonomische Unternehmensperformance bestätigen konnten. Es kann vermutet werden, dass durch den normativen Druck der fokalen Unternehmen die

Unternehmensperformance der Lieferanten verbessert wird, aber nicht unbedingt die des fokalen Unternehmens. LAÜ stellt somit mehr eine Investition in die Lieferantenentwicklung dar (J. Wang & Dai, 2018, S. 14), als dass das fokale Unternehmen selbst ökonomisch davon profitieren kann. Diese Vermutung kann durch die (nicht signifikante) negative Tendenz des Ergebnisses aufgestellt werden, was aber kein Beleg für diese Überlegung darstellt. Eine weitere Erklärung des Ergebnisses geht auf die Überlegungen von Reuter, Foerstl, Hartmann und Blome (2010, S. 56) zurück. Zwar können durch eine strikte Selektierung nicht-ökologisch und nicht-sozial nachhaltige Lieferanten aus dem Portfolio eliminiert werden, jedoch könnte die Reduzierung der ohnehin schon begrenzten Anzahl an Lieferanten den Wettbewerb zwischen diesen negativ beeinflussen, was im Umkehrschluss für das fokale Unternehmen zu höheren Einkaufspreisen führen könnte (Reuter et al., 2010, S. 56). Somit würde sich die Verhandlungsposition des fokalen Unternehmens deutlich verschlechtern. Weiterhin ist die Selektion und die stetige Überprüfung der Lieferanten mit zusätzlichen Transaktionskosten verbunden (Hansen et al., 2011, S. 92).

Es konnte kein signifikant-positiver Effekt von LAÜ auf die soziale Unternehmensperformance beobachtet werden (β = 0,272). Die Hypothese H6c wurde dementsprechend abgelehnt. Auch Gimenez et al. (2012, S. 155) konnten diesen Zusammenhang nicht statistisch belegen. Es kann vermutet werden, dass LAÜ allein nicht zur Verbesserung der sozialen Unternehmensperformance führt. Um dies zu erreichen, sollte der Fokus nicht nur auf der Auswahl und Überwachung von Lieferanten, sondern auch auf der Kollaboration mit diesen liegen (Spence & Bourlakis, 2009, S. 299). Weiterhin könnte der Grund für dieses Ergebnis sein, dass die SSCM Praktik LAÜ nach nachhaltigen Aspekten nicht ausreichend an die Stakeholder kommuniziert wird (z. B. anhand eines CSR-Reportings) und somit keinen positiven Einfluss auf die Reputation des fokalen Unternehmens hat. Eine weitere Erklärung geht auf die Ergebnisse von Koplin, Seuring und Mesterharm (2007, S. 1058) zurück. Die Autor*innen stellten fest, dass in der deutschen Automobilbranche sozialspezifische Aspekte bei der Auswahl von Lieferanten keine bedeutende Beachtung finden.

Zusammenarbeit mit Lieferanten

Es wurde zwar ein leicht-positiver Effekt von ZL auf die ökologische Unternehmensperformance festgestellt (β = 0,151), welcher sich aber als nicht signifikant herausstellte, wonach entgegen der Erwartungen und der Ergebnisse von z. B. Gimenez et al. (2012, S. 155) Hypothese H7a nicht bestätigt werden konnte. Hier kann vermutet werden, dass die befragten Unternehmen nur bedingt mit ihren Lieferanten zusammenarbeiten, um ökologisch nachhaltiger zu werden. Eventuell bestehen hier Hemmnisse bei den fokalen Unternehmen, ihre Lieferanten in umweltspezifischen Aspekten zu unterstützen. Dies spiegelt sich in den Mittelwerten und Medianen der Indikatoren ZL_3 und ZL_4 wider (vgl. Tabelle B.1 im Anhang B). Die befragten Unternehmen stimmten demnach eher nicht zu, ihre Lieferanten mit finanziellen Mitteln zu unterstützen, damit diese in umweltfreundliche Technologien investieren können (M_{ZL_3} = 2,100; Med_{ZL_3} = 2). Auch das Anbieten von Umweltschulungen für die Mitarbeiter*innen der Lieferanten scheint keine etablierte Praktik zu sein (M_{ZL_4} = 2,100; Med_{ZL_4} = 2).

Gemäß den Erwartungen wurde ein signifikanter und positiver Effekt von ZL auf die ökonomische Unternehmensperformance beobachtet (β = 0,258). Hypothese H7b konnte somit angenommen werden, was mit den Ergebnissen von Gimenez et al. (2012, S. 155) übereinstimmt. Nach der SNT (vgl. Kapitel 2.3.2) kann ein Supply Chain Netzwerk als soziales Netzwerk von miteinander verbundenen Unternehmen verstanden werden, deren Erfolg von interorganisational integrierten Geschäftsprozessen und der kollaborativen Performance der einzelnen Netzwerkteilnehmer abhängt (Varsei et al., 2014, S. 247). Verbunden mit der RBT und NRBV stellen diese Netzwerke wertvolle, seltene und schwer imitierbare Ressourcen dar, die somit die Wettbewerbsfähigkeit des gesamten Netzwerkes erhöhen und somit einen positiven Einfluss auf die ökonomische Unternehmensperformance haben können.

Es konnte ein leicht-positiver, aber nicht signifikanter Einfluss von ZL auf die soziale Unternehmensperformance festgestellt werden (β = 0,147), weshalb H7c verworfen wurde. Dies stimmt zwar mit dem Ergebnis von J. Wang und Dai (2018, S. 14) überein, das Ergebnis von Gimenez et al. (2012, S. 157) zeigte jedoch einen signifikanten und positiven Einfluss. Diese gegensätzlichen Resultate aus der Literatur könnten auf die unterschiedlichen Stichproben zurückzuführen sein. Während J. Wang und Dai (2018, S. 2) eine Studie in chinesischen Unternehmen durchführten, konzentrierten sich Gimenez et al. (2012, S. 153) vorwiegend auf Unternehmen aus westlich geprägten Staaten. Gemäß dem Ergebnis kann vermutet werden, dass in deutschen Unternehmen ZL eher der Verfolgung von ökonomischen Zielen dient, als die ökologische und soziale Unternehmensperformance zu verbessern.

Nachhaltige Beschaffung

Entgegen der Erwartungen wurde kein signifikanter und positiver Einfluss von NB auf die ökologische Unternehmensperformance festgestellt (β = 0,056), sodass H8a verworfen werden musste. Zu diesem Ergebnis kamen auch Green Jr. et al. (2012, S. 298) und Yildiz Çankaya und Sezen (2019, S. 111). Die Autor*innen argumentieren, dass sich NB möglicherweise eher auf die Verbesserung der ökologischen Unternehmensperformance der Lieferanten bezieht, als auf die des fokalen Unternehmens (Yildiz Çankaya & Sezen, 2019, S. 113; Green Jr. et al., 2012, S. 299). Den deskriptiven Statistiken (vgl. Tabelle B.1 im Anhang B) ist zu entnehmen, dass den befragten Unternehmen die Öko-Kennzeichnung bei der Beschaffung von Materialien oder Produkten nicht sehr wichtig zu sein scheint (M_{NB_1} = 2,850; Med_{NB_1} = 3). Somit könnten umweltbelastende Materialien in die Produktionsprozesse des fokalen Unternehmens gelangen, was dazu

führen kann, dass die ökologische Unternehmensperformance nicht verbessert wird.

Auch zwischen NB und der ökonomischen Dimension konnte kein signifikanter Zusammenhang festgestellt werden (β = 0,107), weshalb H8b abgelehnt wurde. Yildiz Çankaya und Sezen (2019, S. 111) kamen ebenfalls zu diesem Ergebnis. Da NB auch als Resultat einer nachhaltigen Lieferantenauswahl gesehen werden kann, lässt sich dieses Ergebnis analog zur Erklärung der Beziehung von LAÜ und der ökonomischen Unternehmensperformance interpretieren. Die Auswahl und die dauerhafte Überprüfung der Lieferanten führen zu Transaktionskosten (Hansen et al., 2011, S. 92), die sich wiederum negativ auf die ökonomische Unternehmensperformance auswirken können. Nach Yildiz Çankaya und Sezen (2019, S. 112) kann eine externe Praktik nur erfolgreich sein, wenn fokale Unternehmen und ihre Lieferanten eine enge Partnerschaft pflegen, um gemeinsam wertvolle Wettbewerbsvorteile zu generieren. Dies ist wiederum auf die SNT und die RBT zurückzuführen. Haben sich die Unternehmen aber zu einem erfolgreichen Netzwerk zusammengeschlossen, so kann dies langfristig einen positiven Einfluss auf die ökonomische Unternehmensperformance haben. Eine weitere Erklärung könnte sein, dass die Beschaffung nachhaltiger Materialien und Produkte mit höheren Kosten einhergeht, da die Produktion dieser durch die Lieferanten mit höheren Kosten verbunden sein könnte, als bei herkömmlichen Materialien.

Es konnte kein signifikanter und positiver Einfluss von NB auf die soziale Unternehmensperformance beobachtet werden (β = -0,118), sodass H8c nicht bestätigt wurde. Da NB eine ökologische SSCM Praktik darstellt, ist die Verbesserung der sozialen Unternehmensperformance höchstens auf indirekten Weg z. B. über die Verbesserung der ökologischen Unternehmensperformance möglich. Dies würde zu einer höheren Unternehmensreputation und zu einem besseren Produktimage führen. Da NB aber auch keinen signifikanten Effekt auf die ökologische Unternehmensperformance hat, kann auch dies nicht zu einem positiven Einfluss auf die soziale Dimension führen. Weiterhin kann vermutet werden, dass die befragten Unternehmen ihre Anstrengungen in NB nicht ausreichend nach außen kommunizieren, sodass sich dies nicht positiv auf die soziale Unternehmensperformance auswirken kann.

Interaktionen zwischen den drei Dimensionen der Unternehmensperformance

Sowohl im Modell der internen (SM1) als auch der externen (SM2) SSCM Praktiken wurde untersucht, ob die ökologische und soziale Unternehmensperformance einen positiven Einfluss auf die ökologische Unternehmensperformance hat. Weder in SM1 (β = 0,044) noch in SM2 (β = 0,009) konnte ein signifikanter Einfluss der ökologischen auf die ökonomische Unternehmensperformance beobachtet werden, sodass entgegen der Erwartungen und der Ergebnisse von Green Jr. et al. (2012, S. 299) sowie J. Wang und Dai (2018, S. 13) die Hypothese H9a in beiden Modellen verworfen wurde. Dies könnte zum einen an den unterschiedlichen Stichproben liegen. Während Green Jr. et al. (2012) eine Befragung in US-Unternehmen durchführten und J. Wang und Dai (2018) chinesische Unternehmen befragten, fand die vorliegende Befragung in deutschen herstellenden und verarbeitenden Unternehmen statt. Demnach scheint in Deutschland die ökologische Unternehmensperformance (noch) keinen direkten Einfluss auf die ökonomische Dimension zu haben. Es kann vermutet werden, dass der Aufwand in Investitionen nachhaltiger Praktiken den Nutzen dieser noch übersteigt. So wirken sich z. B. die Reduzierung der Emissionen, des Energieverbrauchs, des Abfalls oder anderer umweltschädlicher Stoffe (noch) nicht signifikant auf die damit einhergehenden Kostensenkungen aus.

Erwartungsgemäß konnte sowohl in SM1 (β = 0,610) als auch in SM2 (β = 0,570) ein signifikanter und positiver Zusammenhang zwischen der sozialen und der ökonomischen Unternehmensperformance festgestellt werden, sodass H9b angenommen wurde. Dieses Ergebnis deckt sich mit dem von J. Wang und Dai (2018, S. 13). Ein hohes Level an sozialer Unternehmensperformance könnte als immaterieller und wertvoller Vermögensgegenstand angesehen werden, wodurch gemäß der RBT und NRBV Wettbewerbsvorteile generiert werden können J. Wang und Dai (2018, S. 14). Eine hohe Unternehmensreputation und ein gutes Produktimage fördert nicht nur die Kundenzufriedenheit, sondern auch die Zufriedenheit aller weiteren internen und externen Stakeholder. Dies kann wiederum zu einer Erhöhung des Umsatzes und des Marktanteils führen und somit zur Verbesserung der Sozio-Effizienz des fokalen Unternehmens beitragen. Weiterhin kann anhand der SNT argumentiert werden, dass sich diese positive Entwicklung nicht nur auf das fokale Unternehmen beschränkt, sondern auch die vor- und nachgelagerten Supply Chain Netzwerk Teilnehmer davon profitieren.

5.3. Implikationen für die Praxis

Sowohl das Modell für interne (SM1) als auch das für externe (SM2) SSCM Praktiken weist eine annehmbare Erklärungs- und Prognoserelevanz auf (vgl. Kapitel 4.2). So bringen die Ergebnisse der vorliegenden Arbeit einige wichtige Erkenntnisse für herstellende und verarbeitende Unternehmen in Deutschland und insbesondere deren SCM hervor. Unternehmen erhalten einige Einblicke, wie sie durch die Umsetzung von SSCM Praktiken ihre Unternehmensperformance gemäß der TBL verbessern können.

Wie die Ergebnisse zeigen, können interne SSCM Praktiken wie IU die ökologische und soziale Unternehmensperformance positiv beeinflussen. Zwar sind für die Implementierung einige Investitionen nötig, welche kurzfristig die ökonomische Unternehmensperformance negativ beeinträchtigen. Dennoch sollten die Unternehmen auch auf lange Sicht planen. Maßnahmen wie die Zertifizierung nach der Umweltnormreihe ISO 14000, die strikte Einhaltung von Umweltschutzrichtlinien und eine gezielte Schulung der Mitarbeiter zum Umweltschutz können indirekt dazu beitragen, dass auch die ökonomische Unternehmensperformance langfristig positiv beeinflusst wird. Weiterhin sollten Unternehmen

in NPP investieren, um den Ressourcenverbrauch und die Emissionen durch die Produktionsprozesse zu reduzieren. Dies führt nicht nur zu einer geringeren Schadstoffbelastung für die Mitarbeiter*innen insbesondere in der Produktion, sondern auch für die umliegenden Gemeinden und Regionen der Unternehmen. Die Wiederverwertung von gebrauchten, überschüssigen oder defekten Materialien (IR) scheint für herstellende und verarbeitende Unternehmen in Deutschland zurzeit keine bedeutende SSCM Praktik zu sein. Hier sollten Unternehmen vermehrt in Recyclingtechnologien investieren, um einen Beitrag zur Vermeidung von Verschwendung zu leisten, was wiederum die ökologische und ökonomische Dimension der TBL positiv beeinflusst. Außerdem sollten Unternehmen verstärkt sozialverantwortliche Maßnahmen implementieren. Durch Investitionen in ein gesundes und sicheres Arbeitsumfeld, umfassende Weiterbildungsmöglichkeiten oder flexible Arbeitszeitmodelle steigt die Arbeitsmoral der Mitarbeiter*innen und erhöht somit langfristig die Produktivität des Unternehmens. Die Unterstützung sozialer Projekte, wie z. B. in Bildung, Sport und Kultu,r bereitet den Unternehmen ein hohes Ansehen in ihrer Region, führt zu einer Erhöhung der Unternehmensreputation und verbessert das Image der Produkte des Unternehmens, was im Endeffekt zu einer Verbesserung der sozialen und ökonomischen Unternehmensperformance führt.

Es kann anhand der Ergebnisse konstatiert werden, dass externe SSCM Praktiken in herstellenden und verarbeitenden Unternehmen in Deutschland weniger implementiert sind als interne SSCM Praktiken (vgl. Kapitel 4.2). Doch auch externe SSCM Praktiken, wie eine nachhaltige Distribution (GD) und Beschaffung (NB), die Auswahl und die Überwachung der Lieferanten nach umwelt- und sozialspezifischen Aspekten (LAÜ) oder auch die Zusammenarbeit mit den Lieferanten (ZL) sind wichtige Maßnahmen, um nicht nur die TBL des fokalen Unternehmens zu verbessern, sondern auch die des gesamten Supply Chain Netzwerkes. Schon die Beschaffung von umweltfreundlichen Rohstoffen ist für einen nachhaltigen Produktionsprozess unausweichlich. Unternehmen sollten deshalb viel Aufwand in eine kritische Lieferantenauswahl und regelmäßige Evaluierung stecken, um gewährleisten zu können, dass nur Rohstoffe bzw. Materialien in den Produktionsprozess gelangen, die den eigenen umwelt- und sozialspezifischen Ansprüchen genügen. Kriterien zur Lieferantenauswahl können z. B. die Zertifizierung der Lieferanten nach ISO Normreihe 14000 oder SA 8000 sein. Weiterhin sollten Unternehmen nicht nur auf umwelt- und sozialspezifische Aspekte bei der Auswahl direkter Lieferanten achten, sondern auch auf die der weiter vorgelagerten Tier-2 Lieferanten. Die Ergebnisse zeigen hier, dass die befragten Unternehmen eher weniger die Tier-2 Lieferanten bei der Auswahl der direkten Lieferanten berücksichtigen. Somit sollte eine Lebenszyklusanalyse der Endprodukte des fokalen Unternehmens schon bei der Herstellung des ersten Rohstoffes ansetzen, anstatt erst innerhalb der eigenen Unternehmensgrenzen. Nach der erfolgreichen Auswahl von geeigneten Lieferanten sollten Unternehmen eine langfristige Partnerschaft anstreben. Dies verhindert zum einen die Entstehung unnötiger Transaktions- und Wechselkosten. Zum anderen können gemäß der SNT und der RBT die Unternehmen innerhalb des Supply Chain Netzwerks gegenseitig voneinander profitieren, indem sie wertvolle Ressourcen, wie z. B. Knowhow zu neuen Umwelttechnologien aber auch finanzielle Unterstützung zur Adaption von SSCM Praktiken miteinander austauschen. Dadurch können Wettbewerbsvorteile zu anderen Unternehmen der Branche generiert werden, was wiederum dazu führen kann, dass die TBL des fokalen Unternehmens und zumindest auch die der Upstream-Seite des Supply Chain Netzwerkes auf lange Sicht verbessert wird.

5.4. Limitationen und zukünftiges Forschungspotenzial

Die vorliegende Arbeit unterliegt einigen Limitationen, aus denen sich Potential für zukünftige Forschung bietet. Als erste Limitation ist die niedrige Rücklaufquote bei der Befragung zu nennen. So ist diese im Vergleich zur einschlägigen Literatur mit 12,20 % vergleichsweise gering. Nach Melnyk, Page, Wu und Burns (2012, S. 37) ist eine Rücklaufquote bei Befragungen in der SCM-Forschung von durchschnittlich ca. 33 % gewöhnlich. Da die Befragung aber während der COVID-19-Pandemie durchgeführt wurde (World Health Organization, 2020) und viele der potenziell Befragten sich deshalb in Kurzarbeit oder im Home Office befanden, kann die Rücklaufquote als akzeptabel angesehen werden. Eine Wiederholung der Befragung zu einem günstigeren Zeitpunkt könnte jedoch zu einer höheren Rücklaufquote führen, wodurch die Generalisierbarkeit der Ergebnisse verbessert würde.

Zweitens konnte aufgrund der vergleichsweise niedrigen absoluten Anzahl an vollständig ausgefüllten Fragebögen (N = 61) das theoretisch hergeleitete Forschungsmodell (vgl. Abbildung 3) nicht wie geplant zur Datenanalyse mit der PLS-SGM herangezogen werden, da sonst die Daumenregel von Barclay et al. (1995, S. 292) für die Mindestanzahl an Beobachtungen für eine PLS-SGM nicht eingehalten werden konnte. Deshalb wurde das Forschungsmodell in zwei Submodelle unterteilt (vgl. Anhang C), sodass der Einfluss interner und externer SSCM Praktiken auf die TBL getrennt voneinander betrachtet wurde. Da aber auch interne und externe SSCM Praktiken einen gemeinsamen Einfluss auf die TBL haben könnten, sollte in einer zukünftigen Studie mit einer höheren Anzahl an Beobachtungen (mindestens N = 80) das gesamte theoretische Forschungsmodell anhand der PLS-SGM analysiert und interpretiert werden.

Drittens konnte aufgrund der geringen Anzahl an Beobachtungen nicht anhand einer Multi-Group-Analysis (MGA) überprüft werden, ob es signifikante Unterschiede in den Modellen zwischen Unternehmen verschiedener Größe oder Branchen gibt. Um eine MGA durchführen zu können, muss ebenfalls die Daumenregel von Barclay et al. (1995, S. 292) zur Mindestanzahl an Beobachtungen in jedem einzelnen Modell der MGA erfüllt sein, um valide Ergebnisse erzielen zu können. Da in der vorliegenden Untersuchung mit einer Anzahl an Beobachtungen von N = 61 diese Daumenregel für beide Submodelle nur knapp erfüllt wurde, war es methodisch nicht möglich, eine MGA durchzuführen. Es

wurde lediglich überprüft, ob die Unternehmensgröße als Kontrollvariable einen signifikanten Einfluss auf die endogenen latenten Konstrukte der TBL hat. Demnach wäre die Durchführung einer MGA für zukünftige Untersuchungen interessant, um Unterschiede bei der Adaption von SSCM Praktiken und deren Einfluss auf die TBL in unterschiedlich großen Unternehmen oder verschiedenen Branchen zu analysieren.

Die vierte Limitation bezieht sich auf das Forschungsdesign und die Methodik. Da eine Querschnittsstudie durchgeführt wurde, konnte nur eine Momentaufnahme des aktuellen Stands der Adaption von SSCM Praktiken in den befragten Unternehmen erzeugt werden. Dies kann vor allem bei den Unternehmen zu Verzerrungen in den Antworten führen, die SSCM Praktiken erst kurz vor der Befragung eingeführt und somit noch keine Erfahrungswerte zum Einfluss der Praktiken auf die TBL des Unternehmens haben. Im Rahmen einer Längsschnittstudie oder Paneldatenforschung könnte untersucht werden, wie sich die Adaption der SSCM Praktiken in ein paar Jahren bzw. im Zeitverlauf auf die drei Dimensionen der Unternehmensperformance auswirken (Esfahbodi et al., 2017, S. 26). Als Vergleichsgrundlage für diese weiteren Untersuchungen könnte die vorliegende Arbeit dienen.

Eine weitere Limitation in Bezug auf das Forschungsdesign ist, dass kein Pretest des Fragebogens stattgefunden hat. Es wurden zwar ausschließlich etablierte Konstrukte und Indikatoren aus einschlägiger Fachliteratur entnommen. Da die Untersuchungen der Literatur aber meist im asiatischen Raum und in Entwicklungsländern stattfanden, waren möglicherweise nicht alle Indikatoren auf den deutschen Kontext übertragbar. In einer zukünftigen Untersuchung sollte deshalb ein Pretest mit einigen Praxisexperten aus dem Bereich SCM durchgeführt werden, um zu prüfen, dass alle für den deutschen Kontext wichtigen Facetten in Bezug auf SSCM Praktiken und der Unternehmensperformance durch den Fragebogen erklärt werden.

Weiterhin wurden im Rahmen der vorliegenden Arbeit lediglich SSCM Praktiken in Bezug auf das fokale Unternehmen und dessen Upstream-Stakeholder einbezogen. SSCM Praktiken mit Bezug auf die nachgelagerten Stufen, wie z. B. die Kooperation mit Kunden, wurden bei der Untersuchung nicht berücksichtigt. In einer weiteren Studie könnte das Forschungsmodell um weitere Konstrukte mit Bezug auf nachhaltige Kollaboration mit Kunden des fokalen Unternehmens erweitert werden.

Die letzte Limitation bezieht sich auf das Datenanalyseverfahren. Ein PLS-Strukturmodell muss Rekursivität aufweisen, „d. h. es darf von keiner latenten Variablen eine Kette von Pfeilen ausgehen, die direkt oder indirekt über andere latente Variablen wieder auf die ursprüngliche latente Variable zeigt“ (Henseler, 2005, S. 71). Somit ist es in den vorliegenden Forschungsmodellen nicht möglich gewesen, z. B. Wechselwirkungen zwischen den drei Dimensionen der Unternehmensperformance darzustellen. Wurde in der vorliegenden Untersuchung nur der direkte Einfluss der ökologischen und sozialen auf die ökonomische Unternehmensperformance analysiert, könnte in einer weiteren Untersuchung überprüft werden, ob die ökonomische Dimension auch Effekte auf die ökologische und soziale Unternehmensperformance aufweist.

6. Zusammenfassung und Fazit

Im Folgenden werden die wichtigsten Ergebnisse und die Vorgehensweise der vorliegenden Arbeit in Hinblick auf die Beantwortung der Forschungsfrage zusammengefasst. Anschließend wird eine Einordnung in den aktuellen Forschungsstand vorgenommen und basierend auf Kapitel 5.4 ein Ausblick für zukünftige Forschung gegeben. Das Ziel der vorliegenden Arbeit war es empirisch zu überprüfen, welchen Einfluss das Konzept des SSCM auf die drei Dimensionen der Unternehmensperformance hat. Zur Erreichung des Ziels wurden die folgenden drei Forschungsfragen aufgestellt:

FF1: *Wie lässt sich das Konzept des SSCM definieren?*
FF2: *Welche SSCM Praktiken können aus der aktuellen Fachliteratur identifiziert werden und wie lassen sie sich klassifizieren?*
FF3: *Wie und in welchem Maße wirken sich die identifizierten SSCM Praktiken auf die ökologische, ökonomische und soziale Performance herstellender und verarbeitender Unternehmen in Deutschland aus?*

Um die Forschungsfragen beantworten zu können, wurde zu Beginn ein Verständnis für den Begriff der Nachhaltigkeit entwickelt, da der Nachhaltigkeitsbegriff ein grundlegender Baustein des SSCM-Konzepts ist. Ausgehend vom Ursprung des Begriffs im frühen 18. Jahrhundert, gewann die Nachhaltigkeitsdebatte im Zuge der Klimaerwärmung besonders ab den 1970er Jahren an Bedeutung. Da besonders Industrieunternehmen einen großen Anteil zu weltweiten THG Emissionen oder sozialem Schaden beitragen, wurde es für sie unumgänglich ein Nachhaltigkeitsmanagement einzuführen, welches die simultane Berücksichtigung aller drei Dimensionen der TBL gewährleistet. Der zweite grundlegende Baustein des SSCM-Konzepts ist das SCM. Da in der Literatur eine Vielzahl an SCM-Definitionen existiert, wurde eine für die vorliegende Arbeit gültige Definition festgelegt. Diese richtet sich nach Hahn (1999, S. 851), da diese zum einen die Supply Chain als Netzwerk von Unternehmen beschreibt, was mit der zuvor festgelegten Definition der Supply Chain nach Mentzer et al. (2001, S. 4) einhergeht, und zum anderen neben der Erreichung ökonomischer Ziele auch auf sozio-ökologische Ziele eingeht. Im Anschluss wurden der Nachhaltigkeitsbegriff und das SCM zum SSCM zusammengefügt.

In Bezug auf **FF1** ist das SSCM grundsätzlich die Verbindung des unternehmerischen Nachhaltigkeitsmanagements mit dem SCM. Da auch hier in der Literatur keine Einstimmigkeit zur Definition von SSCM herrscht, wurden verschiedene Definitionen diskutiert. Hierbei wurde die Definition von Seuring und Müller (2008, S. 1700) für die vorliegende Arbeit festgelegt, da hier besonders die langfristige und gleichberechtigte Verbesserung aller drei Dimensionen der TBL des fokalen Unternehmens und dessen Supply Chain Netzwerk im Mittelpunkt steht. Anschließend wurden anhand

der Organisationstheorien RBT, Institutional Theory, Stakeholder Theory und SNT Treiber für die Implementierung von SSCM Praktiken identifiziert. Es konnte festgestellt werden, dass die Motivation zur Umsetzung von SSCM Praktiken sowohl von internen als auch von externen Quellen getrieben wird. In Kapitel 2.4 wurde schließlich der Zusammenhang von SSCM zur Unternehmensperformance hergestellt. Hier wurden anhand einschlägiger Literatur etablierte interne und externe SSCM Praktiken identifiziert und deren Einfluss auf die TBL auf Basis bisheriger Forschung untersucht. Dies diente als Grundlage zur Herleitung von Hypothesen und des theoretischen Forschungsmodells, welches in Hinblick zur Beantwortung der Forschungsfrage die Basis für die durchzuführende Befragung darstellte. Hinsichtlich der Beantwortung von **FF2** konnten als interne SSCM Praktiken das interne Umweltmanagement (IU), nachhaltiges Produkt- und Prozessdesign (NPP), Investment Recovery (IR) sowie soziale Verantwortung gegenüber Mitarbeiter*innen und Gesellschaft (SV) herausgearbeitet werden. Als in der Literatur etablierte externe SSCM Praktiken gelten Green Distribution (GD), Lieferantenauswahl- und Überwachung (LAÜ), Zusammenarbeit mit Lieferanten (ZL) sowie eine nachhaltige Beschaffung (NB). Für alle SSCM Praktiken wurde anhand bisheriger Erkenntnisse aus der Literatur ein positiver Zusammenhang mit der TBL vorhergesagt.

Zur Beantwortung von **FF3** wurde Primärforschung nach der Survey Research Methodik betrieben. Die Stichprobe bestand aus den 500 größten Unternehmen des Abschnitts C nach dem NACE Rev. 2 Code in Deutschland. Der anhand einschlägiger Literatur entwickelte Online-Fragebogen wurde innerhalb einer zweiphasigen, neunwöchigen Befragung an die Unternehmen der Stichprobe über E-Mail und Telefon verteilt, wodurch nach Datenbereinigungen eine effektive Rücklaufquote von 12,20 % (N = 61) erzielt werden konnte. Da das theoretische Forschungsmodell für die zugrundeliegende Anzahl an Beobachtungen zu komplex war, wurde dieses in zwei Submodelle nach internen (SM1) und externen (SM2) SSCM Praktiken unterteilt. Anhand einer PLS-SGM, welche sich aufgrund ihrer Eigenschaften auch für komplexe Modelle mit einer geringen Anzahl an Beobachtungen eignet, wurden anschließend die Messmodelle auf Reliabilität und Validität überprüft. Nach der Elimination weniger invalider Indikatoren stellten sich alle Messmodelle als reliabel und valide heraus, weshalb mit der Evaluation der PLS-Strukturmodelle fortgefahren werden konnte.

In Bezug auf **FF3** waren vor allem die Beziehungen zwischen den exogenen latenten und den endogenen latenten Konstrukten von Interesse. Es konnten einige Beziehungen zwischen SSCM Praktiken und den drei Dimensionen der Unternehmensperformance empirisch belegt werden. Bei der Evaluation des Einflusses von internen SSCM Praktiken auf die TBL wurden insgesamt 6 der 14 Hypothesen bestätigt. Für IU konnten positive Effekte auf die ökologische und soziale Dimension und auf die ökonomische Dimension ein negativer Effekt empirisch nachgewiesen werden. Allerdings wurde auch ein signifikanter und indirekter Effekt von IU über die soziale auf die ökonomische Unternehmensperformance festgestellt. Für NPP wurde lediglich ein positiver Effekt auf die ökonomische Unternehmensperformance belegt. IR konnte keine Effekte auf die TBL vorweisen und SV zeigte statistisch signifikante und positive Effekte auf die ökologische und soziale Dimension der Unternehmensperformance. Weiterhin wurde ein positiver Mediatoreffekt der sozialen Unternehmensperformance zwischen SV und der ökonomischen Unternehmensperformance empirisch belegt. Bei der Evaluation des Einflusses von externen SSCM Praktiken auf die TBL wurden insgesamt 4 der 14 Hypothesen bestätigt. Für GD wurde ein signifikanter und positiver Effekt auf die soziale Dimension der TBL bestätigt. Weiterhin konnte ein indirekter Effekt von GD über den Mediator der sozialen Dimension auf die ökonomische Unternehmensperformance nachgewiesen werden. Einen signifikanten und positiven Effekt auf die ökologische Unternehmensperformance hatte LAÜ. Während für ZL ein positiver Effekt auf die ökonomische Unternehmensperformance bestätigt werden konnte, wurde für NB kein Effekt festgestellt. Weiterhin wurde ein direkter und positiver Effekt der sozialen auf die ökonomische Dimension der TBL bestätigt. Anhand dieser Ergebnisse konnte konstatiert werden, dass interne SSCM Praktiken in deutschen herstellenden und verarbeitenden Unternehmen durchaus etablierter sind, als externe SSCM Praktiken und diese sich deshalb auch stärker auf die TBL der befragten Unternehmen auswirken. Aufbauend auf diesen Erkenntnissen ergaben sich, wie in Kapitel 5.3 ausführlich dargelegt wurde, einige wichtige Implikationen für die Praxis.

Da sich, wie eingangs beschrieben, bisherige Studien der SSCM-Forschung meist auf Unternehmen in Entwicklungsländern oder den asiatischen Raum beschränkten, kann mit den Ergebnissen der vorliegenden Arbeit ein wichtiger Beitrag zur SSCM-Forschung in Deutschland und somit auch für andere Industrienationen mit ähnlichen wirtschaftlichen Strukturen gegeben werden. Während die meisten Autor*innen in ihren Publikationen überwiegend den Einfluss von SSCM Praktiken auf die ökologische und ökonomische Dimension untersuchten, wurde in dieser Arbeit zusätzlich die soziale Dimension in die Untersuchungen mit aufgenommen, wodurch die Erkenntnis erlangt werden konnte, dass die soziale Dimension einen erheblichen positiven Effekt auf die ökonomische Dimension aufweist.

Wie in Kapitel 5.4 bereits ausführlich dargelegt, bietet die vorliegende Arbeit Potenzial für weitere Forschung. Es ist sicherlich sinnvoll, die Untersuchung noch einmal zu einem späteren Zeitpunkt durchzuführen, um einerseits durch eine höhere Anzahl an Beobachtungen die Generalisierbarkeit der Ergebnisse zu verbessern. Andererseits könnten auf Basis dieser Arbeit weitere Untersuchungen anhand einer Längsschnittstudie oder Paneldatenforschung durchgeführt werden, um auch Veränderungen der Ergebnisse im Zeitverlauf überprüfen zu können. Da die vorliegenden Ergebnisse aufgrund methodischer Schwierigkeiten keine Vergleiche zwischen Unternehmen unterschiedlicher Größe oder Branchen zuließen, könnte auch dieser Aspekt für zukünftige SSCM-Forschungen von Interesse sein, wofür der theoretische Rahmen dieser Arbeit als Grundlage dienen könnte.

Mit der anhaltenden Nachhaltigkeitsdebatte stehen besonders Industrieunternehmen unter dem Druck ihrer Stakeholder, da sie einen großen Beitrag zu den weltweiten THG Emissionen und sozialen Schäden beitragen. Um wettbewerbsfähig zu bleiben, sollten sie sich deshalb in der Verantwortung fühlen, proaktiv ihr gesamtes Supply Chain Netzwerk nachhaltig zu gestalten, indem sie interne und externe SSCM Praktiken zur simultanen Verbesserung aller drei Dimensionen der TBL implementieren. Zwar führen nicht alle Investitionen, insbesondere in externe SSCM Praktiken direkt zur Verbesserung der gesamten TBL, doch auf lange Sicht werden sich die Maßnahmen nicht nur ökologisch und sozial, sondern auch ökonomisch rentieren. In Bezug auf die Definition des Nachhaltigkeitsbegriffs der WCED (1987, S. 37) können Unternehmen einen großen Anteil zur globalen Nachhaltigkeit beitragen, solange sie die gegenwärtigen Bedürfnisse aller Stakeholder erfüllen, ohne dass die Erfüllung zukünftiger Stakeholder beeinträchtigt wird.

Literatur

Ageron, B., Gunasekaran, A. & Spalanzani, A. (2012). Sustainable supply management: An empirical study. *International Journal of Production Economics*, *140* (1), 168–182.

Aitken, J. M. (1998). *Supply Chain Integration Within the Context of a Supplier Association: Case Studies of Four Supplier Associations*. Cranfield: Cranfield University.

Albino, V. (2013). Green Economy. In P. Taticchi, P. Carbone & V. Albino (Hrsg.), *Corporate sustainability* (S. 1–25). Berlin Heidelberg: Springer.

Altundas, A., Memeti, M., Rau, J. & Schrag, J. (2015). Erschwert der Kapitalmarkt das Wirtschaften? In U. Sailer (Hrsg.), *Nachhaltige Unternehmensführung - Aktuelle Fragen zur Umsetzung der Nachhaltigkeit* (S. 11–40). Norderstedt: Books on Demand.

Andersen, M. & Skjoett-Larsen, T. (2009). Corporate social responsibility in global supply chains. *Supply Chain Management: An International Journal*, *14* (2), 75–86.

Armstrong, J. S. & Overton, T. S. (1977). Estimating Nonresponse Bias in Mail Surveys. *Journal of Marketing Research*, *14* (3), 396.

Ayres, R., Ferrer, G. & van Leynseele, T. (1997). Eco-efficiency, asset recovery and remanufacturing. *European Management Journal*, *15* (5), 557–574.

Backhaus, K., Erichson, B. & Weiber, R. (2011). *Fortgeschrittene multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Berlin: Springer.

Ball, A. & Craig, R. (2010). Using neo-institutionalism to advance social and environmental accounting. *Critical Perspectives on Accounting*, *21* (4), 283–293.

Barclay, D., Higgins, C. & Thompson, R. (1995). The Partial Least Squares (PLS) Approach to Causal Modelling: Personal Computer Adoption and Use as an Illustration. *Technology Studies*, *2* (2), 285–309.

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, *17* (1), 99–120.

Baumgarten, H. & Darkow, I.-L. (2004). Konzepte im Supply Chain Management. In A. Busch & W. Dangelmaier (Hrsg.), *Integriertes Supply Chain Management - Theorie und Praxis effektiver unternehmensübergreifender Geschäftsprozesse* (S. 91–110). Wiesbaden: Gabler.

Berkhout, P. H., Muskens, J. C. & Velthuijsen, J. W. (2000). Defining the rebound effect. *Energy Policy*, *28* (6), 425–432.

Bliemel, F., Eggert, A., Fassott, G. & Henseler, J. (Hrsg.). (2005). *Handbuch PLS-Pfadmodellierung: Methode, Anwendung, Praxisbeispiele*. Stuttgart: Schäffer-Poeschel Verlag.

BMU. (2018). *Klimaschutz in Zahlen: Fakten, Trends und Impulse deutscher Klimapolitik*. Berlin: Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit (BMU).

Boone, T., Jayaraman, V. & Ganeshan, R. (2012). *Sustainable supply chains - Models, methods, and public policy implications*. New York: Springer.

Bretzke, W.-R. & Barkawi, K. (2012). *Nachhaltige Logistik: Antworten auf eine globale Herausforderung* (2. Aufl. Aufl.). Berlin Heidelberg: Springer.

Brüssel, C. (2018). Kernkompetenz Nachhaltigkeit und Corporate Social Responsibility. In S. Brüggemann, C. Brüssel & D. Härthe (Hrsg.), *Nachhaltigkeit in der Unternehmenspraxis - Impulse für Wirtschaft und Politik* (S. 11–24). Wiesbaden: Springer Fachmedien.

Busch, A. & Dangelmaier, W. (Hrsg.). (2004). *Integriertes Supply Chain Management - Theorie und Praxis effektiver unternehmensübergreifender Geschäftsprozesse* (2. Aufl. Aufl.). Wiesbaden: Gabler.

Busse, C. (2016). Doing Well by Doing Good? The Self-interest of Buying Firms and Sustainable Supply Chain Management. *Journal of Supply Chain Management*, *52* (2), 28–47.

Busse, C., Schleper, M. C., Niu, M. & Wagner, S. M. (2016). Supplier development for sustainability: contextual barriers in global supply chains. *International Journal of Physical Distribution & Logistics Management*, *46* (5), 442–468.

Campbell, D. T. (1955). The Informant in Quantitative Research. *American Journal of Sociology*, *60* (4), 339–342.

Carnau, P. (2011). *Nachhaltigkeitsethik - Normativer Gestaltungsansatz für eine global zukunftsfähige Entwicklung in Theorie und Praxis* (1. Aufl. Aufl.). München: Rainer Hampp Verlag.

Carter, C. R., Ellram, L. M. & Ready, K. J. (1998). Environmental Purchasing: Benchmarking Our German Counterparts. *International Journal of Purchasing and Materials Management*, *34* (3), 28–38.

Carter, C. R. & Jennings, M. M. (2002). Logistics Social Responsibility: An Integrative Framework. *Journal of Business Logistics*, *23* (1), 145–180.

Carter, C. R., Kale, R. & Grimm, C. M. (2000). Environmental purchasing and firm performance: an empirical investigation. *Transportation Research Part E: Logistics and Transportation Review*, *36* (3), 219–228.

Carter, C. R. & Rogers, D. S. (2008). A framework of sustainable supply chain management: moving toward new theory. *International Journal of Physical Distribution & Logistics Management*, *38* (5), 360–387.

Chabowski, B. R., Mena, J. A. & Gonzalez-Padron, T. L. (2011). The structure of sustainability research in marketing, 1958–2008: a basis for future research opportunities. *Journal of the Academy of Marketing Science*, *39* (1), 55–70.

Chan, R. Y., He, H., Chan, H. K. & Wang, W. Y. (2012). Environmental orientation and corporate performance: The mediation mechanism of green supply chain management and moderating effect of competitive intensity. *Industrial Marketing Management*, *41* (4), 621–630.

Charkovska, N., Halushchak, M., Bun, R., Nahorski, Z., Oda, T., Jonas, M. & Topylko, P. (2019). A high-definition spatially explicit modelling approach for national greenhouse gas emissions from industrial processes: reducing the errors and uncertainties in global emission modelling. *Mitigation and Adaptation Strategies for Global Change*, *24* (6), 907–939.

Chin, W. W. (1998a). Issues and Opinion on Structural Equation Modeling. *MIS Quarterly*, *22* (1), vii–xvi.

Chin, W. W. (1998b). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Hrsg.), *Modern methods for business research* (S. 295–336). London: Psychology Press.

Chopra, S. & Meindl, P. (2013). *Supply chain management: Strategy, planning, and operation* (5. Aufl. Aufl.). Boston: Pearson.

Christopher, M. (2008). *Logistics and supply chain management: Creating value-adding networks* (3. Aufl. Aufl.). Harlow: Financial Times/Prentice Hall.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl. Aufl.). New York: Taylor and Francis. Zugriff auf http://gbv.eblib.com/patron/FullRecord.aspx?p=1192162

Colsman, B. (2013). *Nachhaltigkeitscontrolling - Strategien, Ziele, Umsetzung*. Wiesbaden: Springer Fachmedien.

Cooper, M. C., Lambert, D. M. & Pagh, J. D. (1997). Supply Chain Management: More Than a New Name for Logistics. *The International Journal of Logistics Management*, *8* (1), 1–14.

Corbett, C. J. & Klassen, R. D. (2006). Extending the Horizons: Environmental Excellence as Key to Improving Operations. *Manufacturing & Service Operations Management*, *8* (1), 5–22.

Corsten, H. & Gössinger, R. (2008). *Einführung in das Supply Chain Management* (2., vollst. überarb. und wesentlich erw. Aufl. Aufl.). München: Oldenbourg.

Costanza, R. & Patten, B. C. (1995). Defining and predicting sustainability. *Ecological Economics*, *15* (3), 193–196.

Cruz, J. M. & Wakolbinger, T. (2008). Multiperiod effects of corporate social responsibility on supply chain networks, transaction costs, emissions, and risk. *International Journal of Production Economics*, *116* (1), 61–74.

Cudney, E. & Elrod, C. (2011). A comparative analysis of integrating lean concepts into supply chain management in manufacturing and service industries. *International Journal of Lean Six Sigma*, *2* (1), 5–22.

Dahm, J. D. (2019). *Benchmark Nachhaltigkeit: Sustainability Zeroline - Das Maß für eine zukunftsfähige Ökonomie* (1. Aufl. Aufl.). Bielefeld: transcript.

Das, D. (2017). Development and validation of a scale for measuring Sustainable Supply Chain Management practices and performance. *Journal of Cleaner Production*, *164*, 1344–1362.

Das, D. (2018). Sustainable supply chain management in Indian organisations: an empirical investigation. *International Journal of Production Research*, *56* (17), 5776–5794.

Delmas, M. & Montiel, I. (2009). Greening the Supply Chain: When Is Customer Pressure Effective? *Journal of Economics & Management Strategy*, *18* (1), 171–201.

Delmas, M. & Toffel, M. W. (2004). Stakeholders and environmental management practices: an institutional framework. *Business Strategy and the Environment*, *13* (4), 209–222.

Diesendorf, M. (2000). Models of sustainability and sustainable development. In D. Dunphy, J. Benveniste, A. Griffiths & P. Sutton (Hrsg.), *Sustainability: The corporate challenge of the 21st century* (S. 19–37). Sydney: Allen & Unwin.

DiMaggio, P. J. & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, *48* (2), 147.

Dyckhoff, H. & Ahn, H. (2001). Sicherstellung der Effektivität und Effizienz der Führung als Kernfunktion des Controlling. *Controlling und Management*, *45* (2), 111–121.

Dyllick, T. & Hockerts, K. (2002). Beyond the business case for corporate sustainability. *Business Strategy and the Environment*, *11* (2), 130–141.

Ehrgott, M., Reimann, F., Kaufmann, L. & Carter, C. R. (2013). Environmental Development of Emerging Economy Suppliers: Antecedents and Outcomes. *Journal of Business Logistics*, *34* (2), 131–147.

Elkington, J. (1999). *Cannibals with forks - The triple bottom line of 21st century business*. Oxford: Capstone.

Ellram, L. M. (1990). The Supplier Selection Decision in Strategic Partnerships. *Journal of Purchasing and Materials Management*, *26* (4), 8–14.

Esfahbodi, A., Zhang, Y., Watson, G. & Zhang, T. (2017). Governance pressures and performance outcomes of sustainable supply chain management – An empirical analysis of UK manufacturing industry. *Journal of Cleaner Production (Accepted Manuscript)*, *155* (2), 1–35.

Europäische Kommission. (2008). *NACE Rev. 2 - Statistische Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft*. Luxemburg: Amt für Veröffentlichungen der Europäischen Union.

Fassott, G. & Eggert, A. (2005). Zur Verwendung formativer und reflektiver Indikatoren in Strukturgleichungsmodellen: Bestandsaufnahme und Anwendungsempfehlungen. In F. Bliemel, A. Eggert, G. Fassott & J. Henseler (Hrsg.), *Handbuch PLS-Pfadmodellierung: Methode, Anwendung, Praxisbeispiele* (S. 31–47). Stuttgart: Schäffer-Poeschel Verlag.

Faulbaum, F. (2019). *Methodische Grundlagen der Umfrageforschung*. Wiesbaden: VS Verlag Sozialwissenschaften.

Fischer, J. & Städler, M. (1999). Efficient Consumer Response und zwischenbetriebliche Integration. In H. Hippner, M. Meyer & K. D. Wilde (Hrsg.), *Computer Based Marketing: Das Handbuch zur Marketinginformatik* (S. 349–356). Wiesbaden: Vieweg+Teubner Verlag.

Florida, R. (1996). Lean and Green: The Move to Environmentally Conscious Manufacturing. *California Management Review*, *39* (1), 80–105.

Fornell, C. & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18* (1), 39.

Fowler, F. J. (1988). *Survey research methods* (2. überarbeitete Aufl. Aufl.). Newbury Park: Sage.

Freeman, R. E. (2010). *Strategic management: A stakeholder approach*. Cambridge: Cambridge University Press.

Gallopín, G. C. (2003). *A systems approach to sustainability and sustainable development*. Santiago de Chile: ECLAC.

Gavronski, I., Klassen, R. D., Vachon, S. & Nascimento, L. F. M. d. (2011). A resource-based view of green supply management. *Transportation Research Part E: Logistics and Transportation Review*, *47* (6), 872–885.

Giddings, B., Hopwood, B. & O'Brien, G. (2002). Environment, economy and society: fitting them together into sustainable development. *Sustainable Development*, *10* (4), 187–196.

Gimenez, C., Sierra, V. & Rodon, J. (2012). Sustainable operations: Their impact on the triple bottom line. *International Journal of Production Economics*, *140* (1), 149–159.

Gogoll, F. & Wenke, M. (2017). *Unternehmensethik, Nachhaltigkeit und Corporate Social Responsibility - Instrumente zur Einführung eines Verantwortungsmanagements in Unternehmen* (1. Aufl. Aufl.). Stuttgart: Verlag W. Kohlhammer.

Gold, S., Seuring, S. & Beske, P. (2010). Sustainable supply chain management and inter-organizational resources: a literature review. *Corporate Social Responsibility and Environmental Management*, *17* (4), 230–245.

Golicic, S. L. & Smith, C. D. (2013). A Meta–Analysis of Environmentally Sustainable Supply Chain Management Practices and Firm Performance. *Journal of Supply Chain Management*, *49* (2), 78–95.

Göpfert, I. (2004). Einführung, Abgrenzung und Weiterentwicklung des Supply Chain Managements. In A. Busch & W. Dangelmaier (Hrsg.), *Integriertes Supply Chain Management - Theorie und Praxis effektiver unternehmensübergreifender Geschäftsprozesse* (S. 25–45). Wiesbaden: Gabler.

Green Jr., K. W., Zelbst, P. J., Meacham, J. & Bhadauria, V. S. (2012). Green supply chain management practices: impact on performance. *Supply Chain Management: An International Journal*, *17* (3), 290–305.

Grün, O. & Brunner, J.-C. (2008). Das Aufgabenfeld der Beschaffung. In S. Kummer, O. Grün & W. Jammernegg (Hrsg.), *Grundzüge der Beschaffung, Produktion und Logistik* (S. 62–71). München: Pearson Studium.

Gualandris, J. & Kalchschmidt, M. (2014). Customer pressure and innovativeness: Their role in sustainable supply chain management. *Journal of Purchasing and Supply Management*, *20* (2), 92–103.

Guang Shi, V., Lenny Koh, S. C., Baldwin, J. & Cucchiella, F. (2012). Natural resource based green supply chain management. *Supply Chain Management: An International Journal*, *17* (1), 54–67.

Hackett, G. (1981). Survey Research Methods. *The Personnel and Guidance Journal*, *59* (9), 599–604.

Hahn, D. (1999). Thesen für die Zukunft des Beschaffungsmanagements in einem integrierten Supply Chain Management. In D. Hahn & L. Kaufmann (Hrsg.), *Handbuch industrielles Beschaffungsmanagement - Internationale Konzepte, innovative Instrumente, aktuelle Praxisbeispiele* (S. 849–855). Wiesbaden: Gabler.

Hair, J. F., Hult, G. T. M., Ringle, C. M. & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2. Aufl. Aufl.). Los Angeles: Sage.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Richter, N. F. & Hauff, S. (2017). *Partial Least Squares Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung*. München: Franz Vahlen.

Hair, J. F., Ringle, C. M. & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, *19* (2), 139–152.

Handfield, R. B. & Nichols, E. L. (1999). *Introduction to supply chain management*. Upper Saddle River N.J.: Prentice-Hall.

Hansen, E. G., Harms, D. & Schaltegger, S. (2011). Sustainable Supply Chain Management im globalen Kontext - Praxisstand des Lieferantenmanagements in DAX- und MDAX-Unternehmen. *Die Unternehmung*, *65* (2), 81–110.

Hart, S. L. (1995). A Natural-Resource-Based View of the Firm. *Academy of Management Review*, *20* (4), 986–1014.

Henseler, J. (2005). Einführung in die PLS-Pfadmodellierung. *Wirtschaftswissenschaftliches Studium*, *34* (2), 70–75.

Henseler, J., Ringle, C. M. & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43* (1), 115–135.

Ho, J. C., Shalishali, M. K., Tseng, T.-L. & Ang, D. S. (2009). Opportunities in Green Supply Chain Management. *The Coastal Business Journal*, *8* (1), 18–31.

Hsu, C.-C., Choon Tan, K., Hanim Mohamad Zailani, S. & Jayaraman, V. (2013). Supply chain drivers that foster the development of green initiatives in an emerging economy. *International Journal of Operations & Production Management*, *33* (6), 656–688.

Huang, Y. A., Weber, C. L. & Matthews, H. S. (2009). Categorization of Scope 3 emissions for streamlined enterprise carbon footprinting. *Environmental science & technology*, *43* (22), 8509–8515.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: a review of four recent studies. *Strategic Management Journal*, *20* (2), 195–204.

Hunke, K. & Prause, G. (2014). Sustainable Supply Chain Management in German Automotive Industry: Experiences and Success Factors. *Journal of Security and Sustainability Issues*, *3* (3), 15–22.

IPCC. (2014). *Climate change 2014: Synthesis report - Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva: Autor.

ISO. (2015). *Introduction to ISO 14001:2015*. Geneva: ISO Central Secre-

tariat.

Jarvis, C. B., MacKenzie, S. B. & Podsakoff, P. M. (2003). A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *Journal of Consumer Research*, *30* (2), 199–218.

Kafa, N., Hani, Y. & El Mhamedi, A. (2013). Sustainability Performance Measurement for Green Supply Chain Management. *IFAC Proceedings Volumes*, *46* (24), 71–78.

Kaiser, H. F. & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, *34* (1), 111–117.

Kicherer, A., Schaltegger, S., Tschochohei, H. & Pozo, B. F. (2007). Eco-efficiency. *The International Journal of Life Cycle Assessment*, *12* (7), 537–543.

Klassen, R. D. & Vachon, S. (2003). Collaboration and Evaluation in the Supply Chain: The Impact On Plant-Level Environmental Investment. *Production and Operations Management*, *12* (3), 336–352.

Klassen, R. D. & Whybark, D. C. (1999). The Impact of Environmental Technologies on Manufacturing Performance. *Academy of Management Journal*, *42* (6), 599–615.

Kleindorfer, P. R., Singhal, K. & Wassenhove, L. N. (2005). Sustainable Operations Management. *Production and Operations Management*, *14* (4), 482–492.

Kock, N. (2015). Common Method Bias in PLS-SEM. *International Journal of e-Collaboration*, *11* (4), 1–10.

Koplin, J., Seuring, S. & Mesterharm, M. (2007). Incorporating sustainability into supply management in the automotive industry – the case of the Volkswagen AG. *Journal of Cleaner Production*, *15* (11-12), 1053–1062.

Krause, D. R., Scannell, T. V. & Calantone, R. J. (2000). A Structural Analysis of the Effectiveness of Buying Firms' Strategies to Improve Supplier Performance. *Decision Sciences*, *31* (1), 33–55.

Kurz, R. & Wild, W. (2015). Nachhaltigkeit und Unternehmen. *Umweltwirtschaftsforum*, *23* (4), 323–328.

Lambert, D. M., Cooper, M. C. & Pagh, J. D. (1998). Supply Chain Management: Implementation Issues and Research Opportunities. *The International Journal of Logistics Management*, *9* (2), 1–20.

Lee, H. L., Padmanabhan, V. & Whang, S. (1997). Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science*, *43* (4), 546–558.

Linton, J. D., Klassen, R. & Jayaraman, V. (2007). Sustainable supply chains: An introduction. *Journal of Operations Management*, *25* (6), 1075–1082.

Lu, R. X., Lee, P. K. & Cheng, T. (2012). Socially responsible supplier development: Construct development and measurement validation. *International Journal of Production Economics*, *140* (1), 160–167.

Lu, Y., Zhao, C., Xu, L. & Shen, L. (2018). Dual Institutional Pressures, Sustainable Supply Chain Practice and Performance Outcome. *Sustainability*, *10* (9), 1–25.

Malhotra, M. K. & Grover, V. (1998). An assessment of survey research in POM: from constructs to theory. *Journal of Operations Management*, *16* (4), 407–425.

Mann, B. J. S. & Kaur, H. (2020). Sustainable Supply Chain Activities and Financial Performance: An Indian Experience. *Vision: The Journal of Business Perspective*, *24* (1), 60–69.

Matzler, K., Strobl, A., Thurner, N. & Füller, J. (2015). Switching experience, customer satisfaction, and switching costs in the ICT industry. *Journal of Service Management*, *26* (1), 117–136.

McGraw, D. L. & Watson, G. L. (1976). *Political and social inquiry*. New York: Wiley.

Meadows, D. H., Meadows, D. L., Randers, J. & Behrens, W. W. (1972). *The limits to growth: A report for the Club of Rome's Project on the Predicament of Mankind* (2. Aufl. Aufl.). New York: Universe Books.

Mebratu, D. (1998). Sustainability and sustainable development. *Environmental Impact Assessment Review*, *18* (6), 493–520.

Melnyk, S. A., Page, T. J., Wu, S. J. & Burns, L. A. (2012). Would you mind completing this survey: Assessing the state of survey research in supply chain management. *Journal of Purchasing & Supply Management*, *18* (1), 35–45.

Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D. & Zacharia, Z. G. (2001). Defining Supply Chain Management. *Journal of Business Logistics*, *22* (2), 1–25.

Millar, C., Hind, P. & Magala, S. (2012). Sustainability and the need for change: organisational change and transformational vision. *Journal of Organizational Change Management*, *25* (4), 489–500.

Min, H. & Galle, W. P. (1997). Green Purchasing Strategies: Trends and Implications. *International Journal of Purchasing and Materials Management*, *33* (2), 10–17.

Mitra, S. & Datta, P. P. (2014). Adoption of green supply chain management practices and their impact on performance: an exploratory study of Indian manufacturing firms. *International Journal of Production Research*, *52* (7), 2085–2107.

Müller-Christ, G. (2010). *Nachhaltiges Management - Einführung in Ressourcenorientierung und widersprüchliche Managementrationalitäten* (1. Aufl. Aufl.). Baden-Baden: Nomos.

Ninlawan, C., Seksan, P., Tossapol, K. & Pilada, W. (2010). *The Implementation of Green Supply Chain Management Practices in Electronics Industry*. Hong Kong: Proceedings of the International Multi Conference of Engineers and Computer Scientists 2010, International Association of Engineers.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psycometric theory* (3. Aufl. Aufl.). New York: McGraw-Hill.

OECD. (2006). *Improving Recycling Markets*. Paris: OECD Publishing. Zugriff auf http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=184170 doi: 10.1787/9789264029583-en

Oliver, C. (1991). Strategic Responses to Institutional Processes. *Academy of Management Review*, *16* (1), 145–179.

Pagell, M. & Wu, Z. (2009). Building a More Complete Theory of Sustainable Supply Chain Management Using Case Studies of Ten Exemplars. *Journal of Supply Chain Management*, *45* (2), 37–56.

Panigrahi, S. S., Bahinipati, B. & Jain, V. (2019). Sustainable supply chain management. *Management of Environmental Quality: An International Journal*, *30* (5), 1001–1049.

Paulraj, A. (2011). Understanding the Relationships Between Internal Resources and Capabilities, Sustainable Supply Management and Organizational Sustainability. *Journal of Supply Chain Management*, *47* (1), 19–37.

Paulraj, A., Chen, I. J. & Blome, C. (2017). Motives and Performance Outcomes of Sustainable Supply Chain Management Practices: A Multi-theoretical Perspective. *Journal of Business Ethics*, *145* (2), 239–258.

Peng, D. X. & Lai, F. (2012). Using partial least squares in operations management research: A practical guideline and summary of past research. *Journal of Operations Management*, *30* (6), 467–480.

Petschow, U., Droge, S., Hübner, K. & Meyerhoff, J. (1997). Auswirkungen der Globalisierung auf eine nationale Politik der Nachhaltigkeit. In K. Rennings & O. Hohmeyer (Hrsg.), *Nachhaltigkeit: Nachhaltigkeit und ökologische Ökonomie, Nachhaltigkeit und ökonomische Globalisierung, Nachhaltigkeit und Innovationen* (S. 109–136). Baden-Baden: Nomos Verlagsgesellschaft.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *The Journal of applied psychology*, *88* (5), 879–903.

Podsakoff, P. M. & Organ, D. W. (1986). Self-Reports in Organizational Research: Problems and Prospects. *Journal of Management*, *12* (4), 531–544.

Pokorný, J. & Sokolowsky, P. (1999). A Conceptual Modelling Perspective for Data Warehouses. In M. Nüttgens & A.-W. Scheer (Hrsg.), *Electronic Business Engineering, 4. Internationale Tagung Wirtschaftsinformatik 1999* (S. 665–684). Heidelberg: Physica-Verlag HD. doi: 10.1007/978-3-642-58663-7{\textunderscore}35

Rao, P. & Holt, D. (2005). Do green supply chains lead to competitiveness and economic performance? *International Journal of Operations & Production Management*, *25* (9), 898–916.

Reuter, C., Foerstl, K., Hartmann, E. & Blome, C. (2010). Sustainable Global Supplier Management: The Role of Dynamic Capabilities in Achieving Competitive Advantage. *Journal of Supply Chain Management*, *46* (2), 45–63.

Ringle, C. M., Sarstedt, M. & Straub, D. W. (2012). Editor's Comments: A Critical Look at the Use of PLS-SEM in MIS Quarterly. *MIS Quarterly*, *36* (1), iii–xiv.

Ringle, C. M., Wende, S. & Becker, J.-M. (2020). *SmartPLS3*. Boen-

ningstedt: Smart PLS GmbH. Zugriff am 31.08.2020 auf http://www.smartpls.com

Rivera, J. (2004). Institutional Pressures and Voluntary Environmental Behavior in Developing Countries: Evidence From the Costa Rican Hotel Industry. *Society and Natural Resources*, *17* (9), 779–797.

Roy, T. K., Acharya, R. & Roy, A. (2016). *Statistical Survey Design and Evaluating Impact*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781316550892

Saeed, M. & Kersten, W. (2019). Drivers of Sustainable Supply Chain Management: Identification and Classification. *Sustainability*, *11* (4), 1–23.

Saeed, M., Waseek, I. & Kersten, W. (2017). Literature Review of Drivers of Sustainable Supply Chain Management. In C. Jahn, W. Kersten & C. M. Ringle (Hrsg.), *Digitalization in maritime and sustainable logistics: city logistics, port logistics and sustainable supply chain management in the digital age* (S. 159–184). Berlin: epubli.

Sames, P. C. & Köpke, M. (2012). CO2 Emissions of the Container World Fleet. *Procedia - Social and Behavioral Sciences*, *48*, 1–11.

Sammalisto, K. & Brorson, T. (2008). Training and communication in the implementation of environmental management systems (ISO 14001): a case study at the University of Gävle, Sweden. *Journal of Cleaner Production*, *16* (3), 299–309.

Sarkis, J. (2001). Manufacturing's role in corporate environmental sustainability – Concerns for the new millennium. *International Journal of Operations & Production Management*, *21* (5/6), 666–686.

Sarkis, J. & Talluri, S. (2002). A Model for Strategic Supplier Selection. *Journal of Supply Chain Management*, *38* (1), 18–28.

Sarkis, J., Zhu, Q. & Lai, K.-h. (2011). An organizational theoretic review of green supply chain management literature. *International Journal of Production Economics*, *130* (1), 1–15.

Schaltegger, S. & Harms, D. (2010). *Sustainable Supply Chain Management - Praxisstand in deutschen Unternehmen*. Lüneburg: Centre for Sustainability Management.

Schaltegger, S., Herzig, C., Kleiber, O., Klinke, T. & Müller, J. D. (2007). *Nachhaltigkeitsmanagement in Unternehmen - Von der Idee zur Praxis: Managementansätze zur Umsetzung von Corporate Social Responsibility und Corporate Sustainability*. Lüneburg: Centre for Sustainability Management.

Schaltegger, S., Lüdeke-Freund, F. & Hansen, E. G. (2012). Business cases for sustainability: the role of business model innovation for corporate sustainability. *International Journal of Innovation and Sustainable Development*, *6* (2), 95–119.

Schaltegger, S. & Sturm, A. (1990). Ökologische Rationalität: Ansatzpunkte zur Ausgestaltung von ökologieorientierten Managementinstrumenten. *Die Unternehmung*, *44* (4), 273–290.

Seuring, S. (2013). A review of modeling approaches for sustainable supply chain management. *Decision Support Systems*, *54* (4), 1513–1520.

Seuring, S. & Müller, M. (2008). From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production*, *16* (15), 1699–1710.

Seuring, S. & Müller, M. (2009). Nachhaltiges Management von Wertschöpfungsketten. In A. Baumast (Hrsg.), *Betriebliches Umweltmanagement - Nachhaltiges Wirtschaften in Unternehmen* (S. 165–175). Stuttgart (Hohenheim): Ulmer.

Seuring, S., Sarkis, J., Müller, M. & Rao, P. (2008). Sustainability and supply chain management – An introduction to the special issue. *Journal of Cleaner Production*, *16* (15), 1545–1551.

Shang, K.-C., Lu, C.-S. & Li, S. (2010). A taxonomy of green supply chain management capability among electronics-related manufacturing firms in Taiwan. *Journal of environmental management*, *91* (5), 1218–1226.

Sharma, S. & Gandhi, M. A. (2016). Exploring correlations in components of green supply chain practices and green supply chain performance. *Competitiveness Review*, *26* (3), 332–368.

Shrivastava, P. (1995). The Role of Corporations in Achieving Ecological Sustainability. *Academy of Management Review*, *20* (4), 936–960.

Sikdar, S. K. (2003). Sustainable development and sustainability metrics. *AIChE Journal*, *49* (8), 1928–1932.

Silva, W. H., Guarnieri, P., Carvalho, J. M., Farias, J. S. & Reis, S. A. d. (2019). Sustainable Supply Chain Management: Analyzing the Past to Determine a Research Agenda. *Logistics*, *3* (2), 1–15.

Simchi-Levi, D., Simchi-Levi, E. & Kaminsky, P. (2004). *Managing the supply chain: The definitive guide for the business professional*. New York: McGraw-Hill.

Spence, L. & Bourlakis, M. (2009). The evolution from corporate social responsibility to supply chain responsibility: the case of Waitrose. *Supply Chain Management: An International Journal*, *14* (4), 291–302.

Stahlmann, V. & Clausen, J. (1999). Öko-Effizienz und Öko-Effektivität. *Ökologisches Wirtschaften*, *14* (3), 20–21.

Subramanian, N. & Abdulrahman, M. (2017). An examination of drivers and barriers to reducing carbon emissions in China's manufacturing sector. *The International Journal of Logistics Management*, *28* (4), 1168–1195.

Surroca, J., Tribó, J. A. & Waddock, S. (2010). Corporate responsibility and financial performance: the role of intangible resources. *Strategic Management Journal*, *31* (5), 463–490.

Svensson, G., Ferro, C., Hogevold, N., Padin, C. & Sosa Varela, J. C. (2018). Developing a theory of focal company business sustainability efforts in connection with supply chain stakeholders. *Supply Chain Management: An International Journal*, *23* (1), 16–32.

Sydow, J. (1999). *Strategische Netzwerke - Evolution und Organisation* (1. Aufl., 4. Nachdruck Aufl.). Wiesbaden: Gabler.

Sydow, J. & Windeler, A. (2001). Strategisches Management von Unternehmungsnetzwerken - Komplexität und Reflexivität. In G. Ortmann & J. Sydow (Hrsg.), *Strategie und Strukturation - Strategisches Management von Unternehmen, Netzwerken und Konzernen* (S. 129–143). Wiesbaden: Gabler Verlag.

Thun, J.-H. & Müller, A. (2009). An empirical analysis of green supply chain management in the German automotive industry. *Business Strategy and the Environment*, *19* (2), 119–132.

Vachon, S. & Klassen, R. D. (2006). Extending green practices across the supply chain. *International Journal of Operations & Production Management*, *26* (7), 795–821.

Vachon, S. & Klassen, R. D. (2008). Environmental management and manufacturing performance: The role of collaboration in the supply chain. *International Journal of Production Economics*, *111* (2), 299–315.

Vachon, S. & Mao, Z. (2008). Linking supply chain strength to sustainable development: a country-level analysis. *Journal of Cleaner Production*, *16* (15), 1552–1560.

Vahrenkamp, R., Kotzab, H. & Siepermann, C. (2012). *Logistik: Management und Strategien* (7., überarbeitete und erweiterte Aufl. Aufl.). München: Oldenbourg.

Varsei, M., Soosay, C., Fahimnia, B. & Sarkis, J. (2014). Framing sustainability performance of supply chains with multidimensional indicators. *Supply Chain Management: An International Journal*, *19* (3), 242–257.

von Carlowitz, H. C., Thomasius, H. & Bendix, B. (2013). *Sylvicultura oeconomica - Transkription in das Deutsch der Gegenwart* (Transkription des Originals von 1713 Aufl.). Remagen: Verlag Kessel.

von Hauff, M. (2014). *Nachhaltige Entwicklung - Grundlagen und Umsetzung* (2., aktualisierte Aufl. Aufl.). München: De Gruyter Oldenbourg.

Walton, S. V., Handfield, R. B. & Melnyk, S. A. (1998). The Green Supply Chain: Integrating Suppliers into Environmental Management Processes. *International Journal of Purchasing and Materials Management*, *34* (1), 2–11.

Wang, C. L. & Ahmed, P. K. (2007). Dynamic capabilities: A review and research agenda. *International Journal of Management Reviews*, *9* (1), 31–51.

Wang, J. & Dai, J. (2018). Sustainable supply chain management practices and performance. *Industrial Management & Data Systems*, *118* (1), 2–21.

Wang, Z. & Sarkis, J. (2013). Investigating the relationship of sustainable supply chain management with corporate financial performance. *International Journal of Productivity and Performance Management*, *62* (8), 871–888.

WCED. (1987). *Our Common Future - Report of the World Commission on Environment and Development*. Oxford: University Press.

Weber, M. (2008). The business case for corporate social responsibility: A company-level measurement approach for CSR. *European Management Journal*, *26* (4), 247–261.

Weiber, R. & Mühlhaus, D. (2010). *Strukturgleichungsmodellierung: Eine*

anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS. Heidelberg: Springer.

Welford, R. & Frost, S. (2006). Corporate social responsibility in Asian supply chains. *Corporate Social Responsibility and Environmental Management*, *13* (3), 166–176.

Werner, H. (2017). *Supply Chain Management - Grundlagen, Strategien, Instrumente und Controlling* (6., überarbeitete und aktualisierte Aufl. Aufl.). Wiesbaden: Springer Gabler.

Wilts, H., Lucas, R., von Gries, N. & Zirngiebl, M. (2014). *Recycling in Deutschland – Status quo, Potenziale, Hemmnisse und Lösungsansätze*. Wuppertal: Wuppertal Institut für Klima, Umwelt, Energie GmbH.

WMO. (1979). *Proceedings of the World Climate Conference - A conference of experts on climate and mankind*. Geneva: World Meteorological Organization.

WMO. (2020). *WMO Statement on the State of the Global Climate in 2019*. Geneva: World Meteorological Organization.

Wold, H. (1975). Soft Modelling by Latent Variables: The Non-Linear Iterative Partial Least Squares (NIPALS) Approach. *Journal of Applied Probability*, *12* (S1), 117–142.

World Bank. (1992). *World Development Report 1992: Development and the Environment*. New York: Oxford University Press.

World Health Organization. (2020). *Listings of WHO's response to COVID-19*. Genf. Zugriff am 19.09.2020 auf https://www.who.int/news-room/detail/29-06-2020-covidtimeline

Wu, H.-J. & Dunn, S. C. (1995). Environmentally responsible logistics systems. *International Journal of Physical Distribution & Logistics Management*, *25* (2), 20–38.

Yildiz Çankaya, S. & Sezen, B. (2019). Effects of green supply chain management practices on sustainability performance. *Journal of Manufacturing Technology Management*, *30* (1), 98–121.

Zailani, S., Jeyaraman, K., Vengadasan, G. & Premkumar, R. (2012). Sustainable supply chain management (SSCM) in Malaysia: A survey. *International Journal of Production Economics*, *140* (1), 330–340.

Zhu, Q., Geng, Y. & Lai, K.-h. (2010). Circular economy practices among Chinese manufacturers varying in environmental-oriented supply chain cooperation and the performance implications. *Journal of environmental management*, *91* (6), 1324–1331.

Zhu, Q., Liu, J. & Lai, K.-h. (2016). Corporate social responsibility practices and performance improvement among Chinese national state-owned enterprises. *International Journal of Production Economics*, *171* (3), 417–426.

Zhu, Q. & Sarkis, J. (2004). Relationships between operational practices and performance among early adopters of green supply chain management practices in Chinese manufacturing enterprises. *Journal of Operations Management*, *22* (3), 265–289.

Zhu, Q. & Sarkis, J. (2007). The moderating effects of institutional pressures on emergent green supply chain practices and performance. *International Journal of Production Research*, *45* (18-19), 4333–4355.

Zhu, Q., Sarkis, J. & Geng, Y. (2005). Green supply chain management in China: pressures, practices and performance. *International Journal of Operations & Production Management*, *25* (5), 449–468.

Zhu, Q., Sarkis, J. & Lai, K.-h. (2007). Green supply chain management: pressures, practices and performance within the Chinese automobile industry. *Journal of Cleaner Production*, *15* (11-12), 1041–1052.

Zhu, Q., Sarkis, J. & Lai, K.-h. (2008). Confirmation of a measurement model for green supply chain management practices implementation. *International Journal of Production Economics*, *111* (2), 261–273.

Zijm, H., Klumpp, M., Heragu, S. S. & Regattieri, A. (2019). Operations, Logistics and Supply Chain Management: Definitions and Objectives. In H. Zijm, M. Klumpp, A. Regattieri & S. S. Heragu (Hrsg.), *Operations, Logistics and Supply Chain Management* (S. 27–42). Cham: Springer.

Junior Management Science 7(3) (2022) 802-825

Junior Management Science

journal homepage: www.jums.academy

The Economic Upside of Green Real Estate Investments: Analyzing the Impact of Energy Efficiency on Building Valuation in the Residential Sector

Timo Andreas Deller

Technische Universität München

Abstract

The rising sustainability awareness will affect the carbon-intensive European real estate industry and will force it to adapt to meet climate targets. The purpose of this thesis is to examine whether the energy efficiency of buildings plays a role in the valuation of buildings in the residential sector in the Rhein-Main Region in Germany. This is done by looking at the impact of energy performance certificates of buildings on their rent and sales prices. Data from publicly available real estate advertisements for the years 2019-2020 are analyzed using hedonic regression models. The rent market analysis (N = 44 442) finds significant cold rent premiums of 5.82%, 2.04%, 3.06% for A+, A and B rated buildings compared to the reference level of D. Significant warm rent premiums of 3.86% and 1.98% are found for A+ and B rated buildings. No significant discounts are found for buildings rated below D for cold and warm rents. The sales market analysis (N = 31 426) shows significant premiums of 6.81%, 3.14% and 1.52% for A+, A and B rated buildings, a range of indifference with no premiums or discounts for C to F rated buildings and discounts of -1.73% and -8.80% for G and H rated buildings. The results show that high energy efficiency of buildings creates significant value for investors.

Keywords: Real estate investments; real estate valuation; green buildings; energy efficiency; sustainability.

1. Introduction: Why sustainability matters for the real estate industry

During the last couple of years, climate action has been at the top of the agenda in society, politics and in the economy. Different industries are being shaped by this trend. For example, consumer brands such as Adidas are now selling shoes made out of plastic waste collected from the ocean,[1] the search engine Ecosia supports reforestation around the globe[2] and corporations such as Apple Inc. have published their own targets regarding carbon emissions.[3]

In the field of finance, sustainability aspects have been gaining attention as well. The new field that has emerged is broadly called sustainable finance and comprises all market participant behavior taking sustainability issues into account when making decisions.[4] Further distinctions can be made depending on the objectives of the actors involved. Green Finance, for example, "(...) can be understood as financing of investments that provide environmental benefits in the broader context of environmentally sustainable development."[5] The assets under management in responsible investment funds have increased significantly in the past years in Europe.[6] This reflects the overall increase in demand regarding sustainable assets.

Additional to the increase in demand of sustainable assets, in September 2020 the European Commission announced the plan to adjust the EU climate action target for 2030. The aim to reduce greenhouse emissions will be increased from 40% to 55% compared to the levels of 1990.[7] The climate action targets in general and this additional increase in greenhouse gas reductions by 37,5% are likely to shape future policies and regulations and therefore all industries. The magnitude of the impact this has on different industries depends on their emissions. One very emission in-

[1] Cf. Morgan (2020).
[2] Cf. Eschment (2020).
[3] Cf. Kelion (2020).
[4] Cf. Federal Ministry of Finance (2020).
[5] Ma, Sheren, and Zadek (2016, p. 3).
[6] Cf. KPMG Luxembourg (2019, p. 9).
[7] Cf. European Commission (2020b, p. 1).

DOI: https://doi.org/10.5282/jums/v7i3pp802-825

tensive industry is real estate. According to the 2018 Global Status Report of the Global Alliance for Buildings and Constructions, the real estate industry is responsible for 40% of all energy-related carbon dioxide emissions worldwide.[8] To reduce emissions caused by this sector in Europe, the European Commission has published its Renovation Wave Strategy in October 2020. The aim is to double the current rate of renovation in Europe and increase energy efficiency of buildings.[9] For regulators and industry participants it is crucial to understand the implications of such policy changes. Thus, the current market situation and its response to regulations should be examined. When combining the continuous policy changes with an increased demand for responsible investments, the question arises whether and how energy efficiency changes the valuation of a building. Further, it is relevant to determine whether theoretical adjustments to building valuation can be supported by empirical evidence from real estate market data. Answering the first question lays the theoretical groundwork of this research, while the second question is answered by its empirical findings. The empirical analysis is focused on the regional residential real estate market of the Rhein-Main Region in Germany.[10]

This thesis paper is structured as follows: In section 2, the theoretical approach is discussed: Relevant valuation methods of buildings are introduced, the impact energy efficiency can have on building valuation is shown and a measure for energy efficiency is defined. Section 3 focuses on past research and points to areas where further analysis is needed. Section 4 explains the fundamentals of hedonic price models. In section 5, the data generating process and descriptive statistics are presented. This is followed by the specification of the hedonic models in section 6 and the presentation of the empirical results in section 7. The empirical results are discussed in section 8. The paper concludes with a summary of the main findings and an outlook on future market development and needed research in section 9.

2. Development of the theoretical approach & derivation of hypotheses

2.1. Definition of the thesis perspective, aim, scope and limitations

A sound theoretical approach and study design lay the groundwork for empirical research. Thus, in a first step, the perspective, the aims and the scope and limitations of the study are defined. The perspective analyzed in this paper is that of a real estate investor using equity to invest. Debt financing of real estate as well as mixed financing or other forms of financing are not considered. Financing decisions are assumed to be based on the risk approach of the investor. The factors influencing such decisions are outside the scope of this paper. Further, the research focus lies on single residential building valuation and not portfolio optimization. Additional factors, e.g. diversification aspects to reduce the uncertainty of an investment, come into play when realizing a portfolio of assets.[11] When applying portfolio optimization to real estate, the weight and importance of building characteristics might change. Thus, considering real estate portfolios instead of single buildings in the analysis would skew the measurement of the impact of energy efficiency on the valuation of a single building.

Real estate as an asset is bound to a certain location. Its valuation depends on the local market characteristics.[12] Therefore, the valuation of buildings needs to be a relative comparison between similar assets in the same location. Comparing absolute values between different locations does not appear to be meaningful. The same holds true when talking about premiums or discounts regarding the energy efficiency of buildings or describing other characteristics that are impacting the valuation of the building significantly. Thus, this paper focuses on the local market of the Rhein-Main Region in Germany, which is one of the metropolitan regions in Germany.[13] It is further assumed to be of considerable interest for real estate investors as this region is the financial hub of Germany. This position has been strengthened by the decision and subsequent process of the United Kingdom leaving the EU common market.[14]

Real estate transaction decisions of investors may also depend on taxes. As the focus of this paper is the actual valuation of the building, taxes are a not a part of the analysis. Further, tax laws can change at any given time. Using current tax regulation would make the analysis meaningful only until the next adjustments take place. Including other tax advantages that stem from e.g. the corporate structure used by an investor are beyond the scope of this analysis.

Summary of the scope of this thesis:

- real estate investor perspective with equity only
- single residential building valuation, no portfolio
- findings only applicable for local market in the Rhein-Main Region
- analysis does not consider tax laws

2.2. Summary of the fundamentals of real estate valuation

One of the first researchers to formulate a general theory on how to calculate the value of an investment was John Burr Williams. In his book “The Theory Of Investment Value” he states: “The purchase of a stock or bond, like other transactions which give rise to the phenomenon of interest, represents the exchange of present goods for future goods – dividends, or coupons and principal, in this case being the claim

[8] Cf. International Energy Agency and the United Nations Environment Programme (2018, p. 9).
[9] Cf. European Commission (2020a, p. 1).
[10] Cf. Statista Research Department (2021). The counties and urban districts listed here define this region for the whole thesis.
[11] Cf. Markowitz (1991, p. 470).
[12] Cf. Belke and Keil (2017, p. 17).
[13] Cf. Gesetz über die Metropolregion Frankfurt/Rhein-Main (MetropolG).
[14] Cf. Schleidt (2020).

on future goods. To appraise the investment value, then, it is necessary to estimate the future payments. The annuity of payments, adjusted for changes in the value of money itself, may then be discounted at the pure interest rate demanded by the investor."[15] In other words: the value of the investment today is equal to all future discounted cash flows produced by the asset. Williams applies this to stocks and bonds. The general underlying method of discounting future cashflows can also be applied to other cashflow producing assets such as real estate. How to apply the discounted cash flow (DCF) method to the asset class of real estate is described in detail by Baum and Hartzell (2021).[16] Based on this more recent publication, the relevant elements of this valuation method regarding real estate are summarized in this subsection. Next, it is discussed how energy efficiency can influence them. The theory and methods in this subsection are taken from Baum and Hartzell (2021)[17] if no other source is given. The price of an asset today based on the DCF method is defined by the following equation:

$$Price_0 = \sum_{t=1}^{T} E(CF_t)/(1+r)^t \quad (1)$$

The price in period zero is equal to the appropriate value of the asset. The value of t indicates the time period. This value ranges from t = 1 to T and defines the number of summands. The expected cash flow (CF) of a certain period is defined by E (CF_t). This value is then discounted by 1 plus the discount rate of the investor to the power of the time period to reflect the present value of the future cash flow. Based on this equation, two elements can be determined that are essential for determining the value of the asset for the investor:

- future cash flows
- personal discount rate

The future cash flow depends on two different inputs: income and capital.[18] Based on a combination of current data and forecast data, the aim of an investor is to make the most accurate estimation of both inputs that is possible. To conduct the calculation of the future cash flows in detail, the investment intentions need to be defined. This includes the holding period of the asset. Holding periods of buildings are theoretically unrestricted. What is important to note is that a shorter holding period results in a higher dependency of the return of investment on the sales estimate of the asset. Since the sales estimate is less predictable than rent revenue, risk may be increased with a shorter holding period.[19] After defining the investment intentions, the estimates regarding income and capital inputs are calculated. Factors such as the depreciation of the asset and the occurring expenses[20] need to be considered. The income input is defined by the lease rent. The capital input is the estimation of the resale price.

Relevant for the income input for the DCF calculation is the net operating income (NOI). Figure 1 shows how the NOI is calculated. The lease rent is the gross rental revenue and equal to the overall rent paid by the tenant. Other income is e.g. an additional parking space or storage unit let to the tenant. Together, this is equal to the gross potential income. Deducting the average vacancy rate results in the gross effective income. After the subtraction of the operating expenses, the result is equal to the NOI. Adjustments of the different elements used to calculate the NOI for future periods are based on forecasted data of rental value changes in the local market. The impact of the current lease is higher on the valuation of the building when the terms in the contract are longer and the tenant has rights to renew. Lease events that have a significant impact on the cash flow (e.g. early lease termination by tenant) and their probability need to be estimated.[21]

Gross rental revenue
+ Other income
= Gross potential income
- Vacancy
= Gross effective income
- Operating expenses
= Net operating income

Figure 1: Calculation of the net operating income (NOI)[22]

The estimation of the resale price becomes more difficult with a longer holding period. The most common method to calculate the resale price is the capitalization rate approach. This approach is defined by the following equation:

$$MV_T = NOI_{T+1}/cr_T \quad (2)$$

MV stands for the market value of the building in period T, the time period the building is sold. This market value is equal to the NOI expected in the year following the sale (T+1) divided by the capitalization rate at the time of the sale. The estimate of the NOI in the year T + 1 is based on factors such as rental growth and cost growth. Here, depreciation impacts the rental growth factor. The annual average growth rent rate can be calculated using the following equation:

$$Average\ growth\ rate\ per\ annum = (1+g)/(1+d) \quad (3)$$

[15] Williams (1938, p. 55).
[16] Cf. Baum and Hartzell (2021, pp. 109–157).
[17] Cf. Baum and Hartzell (2021, pp. 109–157).
[17] Cf. Baum and Hartzell (2021, p. 146).
[18] Cf. Baum and Hartzell (2021, p. 148).
[19] Cf. Baum and Hartzell (2021, p. 149).
[20] Some expenses occur regularly and need to be considered when calculating the income input. They can include management cost, repair and maintenance and service costs. Depending on the country, parts of these costs are carried by the tenants. Others occur at the beginning and end of the holding period of the building. The expenses when selling the building need to be subtracted from the sales price to reach a net cash flow estimate of the sales price.
[21] Cf. Baum and Hartzell (2021, p. 149).
[22] Cf. Baum and Hartzell (2021, p. 147).

In this equation, g stands for the rental growth rate of new buildings per annum, while d represents the asset specific depreciation rate. Using this, the NOI for T + 1 can be estimated using the current rent of the building.

The capitalization rate gives the expected return of investing in the building and can be calculated using the following equation:

$$K = RFR_R + i + Rp - (G_R + i - D) \quad (4)$$

K stands for the capitalization rate, RFR_R for the real risk-free rate, i for the expected inflation, Rp for the risk premium, G_R for the real rental growth and D for depreciation. To estimate these factors for time period T is more difficult than the rent estimation.[23] To circumvent this problem, the current capitalization rate applicable for the building is adjusted based on projections regarding the overall market capitalization rate and building specific capitalization rate changes due to depreciation.[24] This concludes the discussion of the most relevant aspects regarding future cash flows. Next, the discount rate is considered briefly.

The discount rate is also called target rate or hurdle rate in a real estate context.[25] It is the summation of the risk-free rate (e.g. interest rate on a three-month U.S. Treasury Bill) and the personal risk premium of the investor. This risk premium is a combination of:

- the property market risk premium
- the sector risk premium
- the location premium
- the asset premium[26]

The asset premium is influenced by the tenant, lease, location and building risk associated with the investment.[27] Combining all, the risk premium on average has a magnitude of around 2-5%.[28]

2.3. Impact of energy efficiency on building valuation & derivation of hypotheses

As described above, the DCF method has two main parts: the cash flow and the discount rate. In the following, both parts are examined separately regarding the effect energy efficiency could have on them. When looking at one specific factor, it is considered, ceteris paribus (c.p.), what happens when increasing or decreasing the energy efficiency. Based on the impact shown, hypotheses are formulated. These hypotheses are subsequently tested in the empirical part of this thesis using datasets from the residential real estate market of the Rhein-Main Region.

First, the cash flow is considered: Cash flow is separated into the income and the capital input. The income input is defined by the NOI. The derivation of the NOI was described in subsection 2.2. The lease structure will not be considered since it varies depending on the parties involved. Including effects of energy efficiency on other income and the vacancy rate are beyond the scope of this thesis.[29] They present interesting topics for future research.

The income input is looked at first: The gross effective income (GEI) is the warm rent paid by the tenant without deducting any costs. This warm rent is divided into the cold rent of the apartment and all the allocable costs. If the energy efficiency of a building is increased, this will result in a decrease in heating costs and thus operating costs. Since these costs are allocable costs, they will reduce the warm rent paid by the tenant, which is a decrease in GEI. The NOI stays the same and therefore the income input does not change. In this scenario, solely the tenant experiences the benefit of higher energy efficiency with the investor remaining indifferent. Such a scenario appears to be unlikely in a rational market environment. Since the tenant's willingness to pay has not changed, it is likely that the tenant is willing to pay the same total expenses, i.e. the same warm rent, for housing. Following this argument, an increase in energy efficiency will lead to an increase of the cold rent while the warm rent stays the same.[30] This scenario means that energy cost savings are capitalized fully into the rent. Whether a 100% capitalization of energy cost savings is possible in all market environments is an interesting question that should be addressed in the future, since the percentage of capitalization may depend on the relative market power of tenant and landlord and may be different between local markets. Answering this question is an opportunity for future research. Going back to the calculation of the NOI, a 100% capitalization means that there is no change in GEI. The operational costs, however, would be decreased. This would lead to a higher cold rent or NOI. As the NOI is the basis for the income input, an investor should value the building higher. Decreasing the level of energy efficiency of the building would lead to the opposite effect. The first hypothesis that can thus be derived from these deliberations is: Net operating income (NOI) is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building and the valuation of the building is increased or decreased respectively.

The arguments above only included capitalization of energy savings and did not require a higher willingness to pay by the tenant, i.e. it is assumed that tenants are indifferent to the energy efficiency of a building and make decisions solely based on warm rent. A question that arises is whether tenants are willing to pay a higher rent for more energy efficient real estate based on the public opinion shift regarding climate change in the past years.[31] This would mean that addi-

[23] Cf. Baum and Hartzell (2021, p. 136).
[24] Cf. Baum and Hartzell (2021, p. 150).
[25] Cf. Baum and Hartzell (2021, p. 153).
[26] Cf. Baum and Hartzell (2021, p. 153).
[27] Cf. Baum and Hartzell (2021, p. 153).
[28] Cf. Baum and Hartzell (2021, p. 154).

[29] Thus, vacancy rate and other income are assumed to be zero. The gross rental revenue equals gross effective income in the remainder of the thesis.
[30] Cf. Cajias and Piazolo (2013, p. 56).
[31] Cf. Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit

tionally to the monetary savings because of a decrease in energy expenses, a utility for the tenant that is derived from the knowledge of living in a more energy efficient building can be identified.[32] Further, it could also be the case that signaling effects influence this market behavior: Tenants may want to show their awareness and personal commitment to other people. This increase in willingness to pay for a more energy efficient building might lead to a higher cold rent (equivalent to NOI).[33] This increase in NOI would not be compensated by a decrease in operating costs. Thus, it would c.p. lead to a higher GEI. The opposite would be true for a less energy efficient building. The tenant's willingness to pay would decrease. The second hypothesis that can be formulated based on this is the following: Gross effective income (GEI) is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building.

Next, the capital input is considered: The capital input is the estimated resale price at the end of the holding period of the building. The question is whether this resale price is also affected by an increase or decrease in energy efficiency of the building. The resale price is determined by the NOI of T+1 and the capitalization rate of T. As the first hypothesis states that an increase/decrease in energy efficiency will increase/decrease the NOI, the same must apply for the resale price when keeping the capitalization rate the same. Based on this, the third hypothesis is: The resale price is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building.

Finally, the discount rate needs to be considered: The discount rate is comprised of the risk-free rate and the personal risk premium of the investor. The risk-free rate cannot be influenced by building characteristics. Therefore, this rate is kept fixed in the following. One part of the personal risk premium of the investor on the other hand is made up of the asset premium. This premium is defined by the building characteristics. The question arises what happens to the asset premium when the energy efficiency of the building is increased/decreased. As public pressure and response from regulators regarding energy efficiency in the real estate sector has increased over the past years, even more extensive changes in regulation are to be expected as we near the emission goals of 2030 and eventually 2050 in the EU and worldwide. Changes such as the ban of oil as heating source may cause a sudden depreciation of assets equipped with oil heating. Whether such an event seems probable to the investor should be reflected in the asset premium and therefore the discount rate. Based on this, it seems reasonable to propose that less energy efficient buildings show an increased asset risk premium.[34] The asset risk premium is also a part of the equation (4) that defines the capitalization rate. Thus, the discount as well as capitalization rate should be higher for less energy-efficient real estate. The opposite should be true for more energy efficient buildings. Combining this with the hypotheses from above, the valuation of the building would increase more than the NOI. This leads to the fourth hypothesis: The market value is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building and the increase/decrease is proportionally bigger than the increase/decrease in NOI.

In hindsight, it needs to be reflected on hypothesis three and four. Hypothesis three considers the estimate of the resale price of a building. This in itself is a building valuation and therefore is hypothesis three similar to hypothesis four. An important difference is the point in time of the two valuations: The resale price is an estimation of the future market value of the building at the end of the holding period. It can only be estimated today. The market value formulated in hypothesis four is the current market value of the building. Should hypothesis four be supported, then hypothesis three would also be supported because a decrease in the discount rate and thus also a decrease of the capitalization rate has taken place. This means that only hypotheses one, two and four need to be examined. In summary, the hypotheses that will be addressed in this paper are:

A. Net operating income (NOI) is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building and the valuation of the building is increased or decreased respectively.

B. Gross effective income (GEI) is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building.

C. The market value is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building and the increase/decrease is proportionally bigger than the increase/decrease in NOI.

2.4. Assessment of the energy performance certificate as indicator for energy efficiency

Defining the energy efficiency of a building in a comparable, reliable, and accurate manner is an important basis for this research paper. Additional aspects that need to be considered are the size of the empirical sample as well as data availability.

Throughout the literature, there is a usage of formal certificates as a basis for such analyses.[35] These certificates differ regarding the focus of examination. There are certificates like the Leadership in Energy and Environmental Design (LEED) and the Building Research Establishment Environmental Assessment Methodology (BREEAM) that consider the overall sustainability of a building.[36] A different certificate, the European Energy Performance Certificate

und Verbraucherschutz and Bundesumweltamt (2020). The significant shift regarding environmental and climate awareness can be seen in 2018 compared to 2016. Thus, for the remainder of this thesis, all data starting with 01/2018 is regarded as capturing this shift in awareness. All data before 01/2018 is seen as not being able to capture this shift in awareness.

[32] Cf. Cajias and Piazolo (2013, p. 189).

[33] Cf. Cajias and Piazolo (2013, p. 56).

[34] Cf. Leopoldsberger, Bienert, Brunauer, Bobsin, and Schützenhofer (2011, p. 117).

[35] Cf. e.g. Fuerst, Oikarinen, and Harjunen (2016, p. 560). Brounen and Kok (2011, p. 169).

[36] Cf. U.S. Green-Towers Sustainable High-Rises GmbH (2021). Building Research Establishment Ltd (2021).

(EPC), focuses exclusively on the energy consumption of the building.[37] The EPC was first introduced in Europe in 2002 by the EU-Directive 2002/91/EC. The legislation regarding the EPC was changed again in 2010 by the EU-Directive 2010/31/EU to make the use of EPCs mandatory in advertisements when selling or leasing a building. In 2018 EU-Directive 2018/844/EU was amended, aiming to increase transparency and consistency of the national calculation methodologies. As this study focuses on data in the Rhein-Main Region in Germany, the respective German EPC, the "Energieausweis", is used.

Comparability: Since the EPC is based on a directive by the European Commission, member states of the EU must pass their own laws regarding its implementation.[38] Thus, data from different EU countries cannot be directly compared or combined into one dataset. Within Germany, however, the EPC is the same making it possible to compare different buildings within this country. Another point of critique is that the average climate of Germany is used for the energy consumption needs calculation.[39] This means that regions above and below the average will have a de facto energy consumption that differs from the one calculated using the rules of the EPC. This limits comparability between regions, at least between those of different climate zones. As we are only looking at data from the Rhein-Main Region, which is assumed to be one climate zone, this aspect can be disregarded.

Reliability: According to the "Gebäudeenergiegesetz" (GEG) and the previously applicable "Energieeinsparverordnung" (EnEV) in Germany, only people with specific training and professional experience are allowed to perform the assessment for an EPC. The GEG lists all requirements such a person has to meet before being considered qualified.[40] Examples include engineers, architects, physicists and craftsmen. No official accreditation is needed to perform an EPC assessment. Based on this, a general reliability between specialists assessing a single object is assumed. Differences in the result caused by people-specific errors cannot be excluded though. For the purpose of this study, it is assumed that the error term in the calculations caused by this human error is evenly spread across the sample. In this case, this error will contribute to the spread of the data but will affect neither average values nor the overall empirical results. Of note, whether this assumption is true cannot be ascertained using the empirical data available.

Accuracy: The calculation performed regarding the energy efficiency of a building is based on the usable floor space and not the living space. Additionally, warm water is sometimes not included if a building does not heat water locally and two methods of calculation within Germany exist that can lead to different values for the same building.[41] These aspects show that the energy consumption value of a building has to be judged in a broader context. This context cannot be explored in this paper because of data limitations. However, this kind of data accuracy may not be needed for the present analysis, as decisions made by the tenant or buyer are likely between similarly constructed and measured buildings.

Data availability: According to the GEG and EnEV, all owners of a residential building must provide an EPC to prospective tenants or buyers.[42] They also must include certain metrics such as the energy consumption amount of kwh / m^2 per annum in a real estate advertisement.[43] Some exceptions to these rules exist. But it can be said that most sale as well as lease offers in the past years should have an EPC available.

In sum, the EPC is a widely used framework allowing the classification of buildings in Germany based on their energy efficiency. The way of calculating the energy-consumption needs of a building has not changed significantly in the past years.[44] This shows the consistency of the certificate. Although there are points that can be criticized, the EPC is the only widely used and available measure of energy efficiency in the residential as well as other building sectors in Germany that could be identified. The EPC is consequently not a perfect but a reasonable proxy measurement for the energy efficiency of a residential building in Germany and can be used as a data source for this research.

3. Review of the extant literature

During the last ca. fifteen years, the research field examining the effects of energy efficiency or - in a broader sense - sustainability of buildings on their sales and rent performance has gained momentum. In this section, the extant literature will be discussed taking into account the following aspects:

- time period analyzed
- location of the real estate market
- building sector
- proxy used to measure energy efficiency
- whether sales or rent prices are considered

In the literature, the sales or rent prices are defined as the dependent variables and analyzed using hedonic regression models.[45]

In the research field of "sustainability and real estate", five comprehensive literature reviews were published recently. Of these five, two have been published in academic

[37] Cf. EU-Directive 2010/31/EU, Article 2, No. 12.

[38] Cf. Communication department of the European Commission (2020).

[39] Cf. Verbraucherzentrale Nordrhein-Westfalen and Verbraucherzentrale Rheinland-Pfalz (2020).

[40] Cf. GEG, §77. EnEV, §21.

[41] Cf. Verbraucherzentrale Nordrhein-Westfalen and Verbraucherzentrale Rheinland-Pfalz (2020).

[42] Cf. GEG, §80. EnEV, §16.

[43] Cf. GEG, §87. EnEV, §16a.

[44] Cf. Verbraucherzentrale Nordrhein-Westfalen and Verbraucherzentrale Rheinland-Pfalz (2020).

[45] Cf. e.g. Wahlström (2016, p. 201). Kholodilin, Mense, and Michelsen (2017, p. 3223). Hyland, Lyons, and Lyons (2013, p. 945). Only a few examples from the academic literature. Others also use this method.

journals,[46] another one is a conference paper[47] and the last two are working papers.[48] All of them support the hypothesis that a price premium for sustainability certificates or energy efficiency of buildings exists.[49] Of note, the magnitude of the price premium significantly depends on the aspects mentioned above like building sector (e.g. commercial or residential) and location.[50] This is the case within as well as between the literature reviews: Ankamah-Yeboah et al. (2014) find a global average price premium of 7,6% for buildings with some form of energy certification.[51] Fizaine et al. (2018) conclude that a premium between 3.5% and 4.5% is present in the literature when controlling for publication bias.[52] Brown and Watkins (2016), only looking at the sales prices in the residential real estate sector, report a mean weighed premium of 4.3%.[53] This is very close to the global premium of 4.2% found by Cespedes-Lopez et al. (2019) for sales prices in the residential real estate sector.[54] As these analyses are based on different studies (15 studies used by Kim et al. (2016)[55] compared to 66 studies used by Cespedes-Lopez et al. (2019)[56]), their data and conclusions cannot be directly compared. This limitation as well as the limitation regarding the accumulated price premiums in the literature reviews was formulated by Cespedes-Lopez et al. (2019) in the most recent literature review published in an academic journal:

"This document is useful in order to understand the current behavior on a global level. However, it has certain limitations due to combining data from distinct studies that are influenced by geographic area, type of qualification used, etc. Therefore, the results should be considered within the context of the analyzed documents and not as evidence of causality."[57]

Thus, it is important to, in a specific manner, define and then identify the space of the relevant academic literature based on the key aspects mentioned above.

This thesis focuses on the most recent (time period) impact of energy efficiency (proxy) on the sales and rent prices (dependent variables) of residential buildings (building sector) in the Rhein-Main Region of Germany (location). Using these parameters as a filter, the body of literature pertinent for this thesis decreases significantly. Three studies published in academic journals remain when relaxing the aspect "time" completely and the aspect "location" from the Rhein-Main Region to all of Germany.[58] The remaining publications all show that a price premium is achieved for more energy efficient buildings.[59] As the context of each study is relevant to understand the findings, a detailed review of these studies is needed to show where additional research can add insights.

Cajias and Piazolo (2013) examined the impact of energy efficiency of residential buildings on their financial performance using data from the German Investment Property Databank (IPD) ranging from the year 2008 until 2010.[60] Of note, this was the time period right after the global economy was hit by the real estate credit crisis. An important aspect of this crisis was investors quickly switching to high-quality and low risk assets in the USA. This led to a significant credit spread of commercial mortgage-backed securities and consequently to a dry up of the loan sector to nearly zero in 2008. The same had generated loans of 230 billion USD in 2007.[61] The consequence of this development was a global capital shortage.[62] The question arises in how far the underlying market conditions of this time period and therefore the findings of the study can be applied to the markets of today. Another time-related aspect of the data is the national standard regarding the levels of energy efficiency in housing. Cajias and Piazolo (2013) state that up to 200kWh, significant rent premiums can be achieved.[63] Since then, the energy classification has been revised. Today, according to the GEG, the G standard of the German EPC starts at 200kWh. This indicates that there has been a significant shift towards more energy efficient buildings that could also have had an impact on the price premium being achieved on the market. The following calculation underscores these considerations: Cajias and Piazolo (2013) state in their paper: "The hedonic results additionally show that one percent energy conservation boosts rent prices by +0.08 percent and market value by +0.45 percent, ceteris paribus."[64] For the housing and the energy standard of today, this statements needs verification, since the difference between a bad performing C rated building (100kWh/(m^2a)) and a bad performing A rated building (50kWh/(m^2a)) is equal to an energy conservation of 50%. Thus, according to Cajias and Piazolo (2013), the market value of the A rated building should be 22.5% higher than that of the C rated building. This estimate seems very high and the question arises whether results based on current data can corroborate these results for present times.

Kholodilin et al. (2017) examine the capitalization of energy savings in rent and sales prices in the regional resi-

[46] Cf. Cespedes-Lopez, Mora-Garcia, Perez-Sanchez, and Perez-Sanchez (2019). Fizaine, Voye, and Baumont (2018) – of note: journal with low impact factor.
[47] Cf. Kim, Lim, and Kim (2016).
[48] Cf. Ankamah-Yeboah, Rehdanz, et al. (2014). Brown and Watkins (2016).
[49] Cf. Ankamah-Yeboah et al. (2014, p. 20). Brown and Watkins (2016, p. 2). Cespedes-Lopez et al. (2019, pp. 53-54). Fizaine et al. (2018, p. 1033). Kim et al. (2016, p. 47).
[50] Cf. Fizaine et al. (2018, p. 1028).
[51] Cf. Ankamah-Yeboah et al. (2014, p.12).
[52] Cf. Fizaine et al. (2018, p. 1017).
[53] Cf. Brown and Watkins (2016, p. 2).
[54] Cf. Cespedes-Lopez et al. (2019, p. 1).
[55] Cf. Kim et al. (2016, p. 43).
[56] Cf. Cespedes-Lopez et al. (2019, p. 1).
[57] Cespedes-Lopez et al. (2019, p. 54).
[58] Cf. Cajias and Piazolo (2013). Kholodilin et al. (2017). Cajias, Fuerst, and Bienert (2019).
[59] Cf. Cajias and Piazolo (2013, p. 53). Kholodilin et al. (2017, p. 3234). Cajias et al. (2019, p. 189).
[60] Cf. Cajias and Piazolo (2013, p. 53).
[61] Cf. Baum and Hartzell (2021, pp. 60-62), for all statements regarding the real estate credit crisis.
[62] Cf. Baum and Hartzell (2021, p. 68).
[63] Cf. Cajias and Piazolo (2013, p. 67).
[64] Cajias and Piazolo (2013, p. 53).

dential real estate market of Berlin.[65] The data used in this publication stem from online housing portals and were collected from June 2011 until December 2014.[66] Although this is a more recent study compared to Cajias and Piazolo (2013), the current opinion shift regarding climate awareness is not yet included in their data. Furthermore, the focus is on Berlin, a regional market in Germany with local regulation limiting the rights of real estate owners: The Kündigungsschutzklausel-Verordnung protects tenants from eviction by owner-occupiers for a period of several years after purchase of the building.[67] The main insight generated by the study is that owner-occupiers and investors in rental buildings capitalize energy savings in sales prices.[68] For a rented building, where energy savings benefit the tenant, the financial savings are a 2.5 multiple of the investor's willingness to pay for a building with such savings, suggesting only a partial capitalization of energy savings in the form of increased rent.[69]

Cajias et al. (2019) examine data from 2013-2017 for all of Germany regarding the impact of energy efficiency (extracted from the EPC) on residential rent prices.[70] This enables them to differentiate between top tier markets such as Munich and Frankfurt and secondary markets. Here, they show that importance of energy efficiency is decreased in the top tier markets, potentially due to high demand and inelastic supply.[71] Another aspect relevant for investors that is introduced by the authors is "time on market" of the buildings. It is found that very energy inefficient buildings remain on the rental market longer.[72] Their study was performed with market data until 2017, which is also the period before the recent change in climate awareness. The question arises, whether we now see a stronger and more distinct premium in the rental market with current data. Further, their study only looks at rent prices. The question how energy efficiency affects the sales price of a building is left unanswered. A third aspect that needs to be considered is the perspective that was taken by the authors: Cajias et al. (2019) looked at the overall German market. They used location control variables and therefore minimized the noise between regions.[73] However, it is not possible to discern region-specific price premiums. Higher price premiums in regions where climate action is a priority for the public could exist compared to other regions where climate action is considered less important. Thus, the empirical results of the overall market analysis are interesting but only of limited use for investors, as investors have to consider local market characteristics and trends. Comparability of their study with the present study is limited since in the present study a region-specific analysis is performed.

The key takeaways from academic literature for the German residential real estate market are:

- rental and sales price premiums for energy efficient buildings are present [74]
- owner-occupiers and investors capitalize energy savings well in sales prices. The magnitude of capitalization, however, is different. This is because energy savings exceed the tenant's willingness to pay by a factor of 2.5 when looking at rent prices [75]
- tight rental markets decrease importance of energy efficiency, which has a greater effect on prices in non-metropolitan regions [76]
- very energy inefficient buildings remain on the rental market longer [77]

A brief expansion of the literature review to the European level seems reasonable since some studies also used the EPC as a proxy for energy efficiency. While doing so, it has to be kept in mind that different national implementations of the EU directive exist, limiting comparability.

Brounen and Kok (2011) performed one of the earlier studies on the residential real estate sector in Europe. Even after controlling for better quality of buildings and thermal characteristics, they found a sales premium of 10.2% for residential buildings with an A energy rating compared to D rated buildings.[78] Similarly, Hyland et al. (2013) in their study on the residential real estate market in Ireland reported a 9% sales premium and a premium of almost 2% in the rental market for properties of energy efficiency level A compared to D.[79] They further find that energy efficiency has a greater impact (almost double) in less liquid markets.[80] In Spain, de Ayala, Galarraga, and Spadaro (2016) find a sales premium of 9.8% for residential buildings that are rated in the A, B or C category compared to the rest.[81] As this is a block comparison, it's magnitude cannot be compared to the results of the other studies. This inconsistency in EPC levels used for comparison was already criticized by Cespedes-Lopez et al. (2019), as it was one of the reasons why the findings of their meta-regression for Europe were inconclusive.[82] Another study that compares the A, B and C rated buildings to the reference level D was done by Fuerst et al. (2016) with data from the residential market in Helsinki, Finland. They find a sales price premium of up to 3.3% for these buildings.[83] The more interesting finding is that

[65] Cf. Kholodilin et al. (2017, p. 3218).
[66] Cf. Kholodilin et al. (2017, p. 3224).
[67] Cf. Kündigungsschutzklausel-Verordnung, §2.
[68] Cf. Kholodilin et al. (2017, pp. 3232–3234).
[69] Cf. Kholodilin et al. (2017, p. 3232).
[70] Cf. Cajias et al. (2019, pp. 177 + 182).
[71] Cf. Cajias et al. (2019, p. 186).
[72] Cf. Cajias et al. (2019, p. 189).
[73] Cf. Cajias et al. (2019, p. 179).
[74] Cf. Cajias and Piazolo (2013, p. 53). Cajias et al. (2019, p. 177).
[75] Cf. Kholodilin et al. (2017, pp. 3232-3234).
[76] Cf. Cajias et al. (2019, p. 189).
[77] Cf. Cajias et al. (2019, p. 189).
[78] Cf. Brounen and Kok (2011, p. 176).
[79] Cf. Hyland et al. (2013, p. 950).
[80] Cf. Hyland et al. (2013, p. 949).
[81] Cf. de Ayala et al. (2016, p. 22).
[82] Cf. Cespedes-Lopez et al. (2019, p. 53).
[83] Cf. Fuerst et al. (2016, p. 567).

even after controlling for neighborhood characteristics and maintenance costs (this includes energy costs), a sales price premium of 1.3% was identified.[84] The authors argue that this premium is evidence for significant signaling effects for energy efficient buildings in the residential real estate market.[85] Contrary findings regarding a premium for lower energy consumption values were found by Wahlström (2016) in Sweden. This author included several building characteristics (e.g. new façade or new roof) having an impact on the energy efficiency. For these characteristics, price premiums were found.[86] For a lower calculated energy consumption itself, no price premium was found.[87] Thus, in Sweden, the buyer values the actual attributes of the building more than its energy consumption needs reflected by the EPC.[88]

Coming back to the market of interest in this study, i.e. the Rhein-Main Region, the question arises how the energy efficiency of buildings is valued in this market and whether significant changes can be identified compared to studies performed with data before 2018. A representative survey published by the Bundesumweltamt in Germany shows that 68% of respondents see environment and climate protection as a very important challenge in 2019 compared to 53% in 2016.[89] Thus, different findings seem plausible. The findings of this analysis will also be compared to Kholodilin et al. (2017) as they looked at another regional market, i.e. Berlin, in Germany. This will help to understand whether the impact of energy efficiency is similar in metropolitan areas.

4. The hedonic price model and its application

In section two, the application of the DCF in a real estate context was discussed. Analyzing how energy efficiency might affect the outcome of this method led to four hypotheses of which three will be further addressed in this thesis:

A. Net operating income (NOI) is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building and the valuation of the building is increased or decreased respectively.

B. Gross effective income (GEI) is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building.

C. The market value is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building and the increase/decrease is proportionally bigger than the increase/decrease in NOI.

Addressing these hypotheses requires analyzing how energy efficiency of buildings is valued in the real estate market of the Rhein-Main Region. Hedonic price models are often used for such an analysis.[90] The statistical methodology employed is a regression analysis and in the case of the present paper a multiple linear regression analysis. Janssen, Söderberg, and Zhou (2001) describe the hedonic model as a framework to analyze goods with a specific set of differentiable characteristics that make up the market value of the good but do not possess a market price on their own. Every characteristic has a certain utility for the user and changes the overall value of the product considered. This includes tangible as well as intangible characteristics.[91]

The earliest applications of the hedonic price model were used to estimate farmland values in Minnesota[92] and Iowa.[93] Another early use resulted in the development of hedonic price indexes for automobiles.[94] Long after their publication, the influence of these early papers on the development of hedonic models and whether Haas (1922) and Wallace (1926) can be regarded as hedonic applications was discussed.[95] These application oriented studies were followed by publications focusing on the consumer[96] and economic[97] theories behind hedonic price models. Lancaster (1966) writes that "(...) consumption is an activity in which goods, singly or in combination, are inputs and in which the output is a collection of characteristics. Utility or preference orderings are assumed to rank collections of characteristics and only to rank collections of goods indirectly through the characteristics that they possess."[98] He viewed the characteristics of the good as source of utility instead of the good in itself. Regarding the prices of these characteristics Rosen (1974) states: "Hedonic prices are defined as the implicit prices of attributes and are revealed to economic agents from observed prices of differentiated products and the specific amounts of characteristics associated with them. They constitute the empirical magnitudes explained by the model. Econometrically, implicit prices are estimated by the first step regression analysis (product price regressed on characteristics) in the construction of hedonic price indexes."[99] Regressing the product price (rent or sales price) on its characteristics will be the basis of the empirical analysis in this paper.

For the estimation of the coefficients of the product characteristics, different methods can be used. In this thesis, the widely used ordinary least squares (OLS) method is applied. It minimizes the squared residuals between estimated and observed values.[100] This method produces the best linear unbiased estimator if the following assumptions are met:

[84] Cf. Fuerst et al. (2016, p. 567).
[85] Cf. Fuerst et al. (2016, p. 560).
[86] Cf. Wahlström (2016, pp. 201-202). Also for the previous sentence.
[87] Cf. Wahlström (2016, pp. 201-202).
[88] Cf. Wahlström (2016, p. 197).
[89] Cf. Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz and Bundesumweltamt (2020).
[90] Cf. e.g. Wahlström (2016, p. 201). Kholodilin et al. (2017, p. 3223). Hyland et al. (2013, p. 945). Only a few examples from the academic literature. Others also use this method.
[91] Cf. Janssen et al. (2001, p. 344).
[92] Cf. Haas (1922, p. 1).
[93] Cf. Wallace (1926, p. 389).
[94] Cf. Leavens (1939, p. 169).
[95] Cf. Colwell and Dilmore (1999, p. 620).
[96] Cf. Lancaster (1966).
[97] Cf. Rosen (1974).
[98] Lancaster (1966, p. 133).
[99] Rosen (1974, p. 34).
[100] Cf. Urban and Mayerl (2018, p. 38).

1. Linearity: The estimated coefficients are of a linear nature.
2. Exogeneity: The mean of the error term is equal to zero.
3. Homoscedasticity: The variance of the error term is constant.
4. Autocorrelation: There exists no covariance between error terms.
5. Multicollinearity: There exists no perfect multicollinearity between explanatory variables.
6. Normality of residuals: The error terms are normally distributed. [101]

Before developing applicable models for analysis, data quality needs to be considered. Data quality is of the essence, since an inappropriate database must by default lead to invalid results and conclusions, regardless of the statistical method employed.

5. Evaluation of the data source & presentation of the descriptive statistics

5.1. Review of the data generating process

Finding pertinent and reliable real estate data for hedonic models is difficult. Real estate brokers are not inclined to give away critical information, including transaction data of brokered deals in the past. Although this problem may be fixable, e.g. by introducing a non-disclosure agreement, this does not solve the second problem: If data from only a small group of real estate brokers is analyzed, a selection bias could be introduced, e.g. observations from real estate brokers with regional specializations or brokers covering only a subsegment of buildings on the market. This could weaken the explanatory power of the model and might lead to a biased estimator. Ideally, an unbiased data set representing the overall market should be analyzed. Kholodilin et al. (2017) collect this information for the German market from online housing portals that have grown significantly in the past.[102] Hyland et al. (2013) collect the data for the Irish residential real estate market from the biggest national housing portal present in Ireland.[103] To increase the comparability between their findings and the findings of this study, a similar approach was used and observations from e.g. major internet portals were used for this analysis. The Real Estate Pilot AG provided the micro-data on the real estate market of the Rhein-Main Region.[104] The time period considered is from January 2019 until December 2020. More specifically, this time period is defined by the date the offer was first seen on the market. The Rhein-Main Region was defined by the counties and cities listed on Statista.[105] The data collected by the Real Estate Pilot AG is extracted from different internet platforms as well as regional and transregional newspapers.[106] A data update is performed by the Real Estate Pilot AG once per day.

Since raw data from the internet are extracted automatically, observations may not always be complete and may even contain erroneous data. Accordingly, others who have used such databases pointed out that the raw data might be biased by duplicated observations, that the online platforms could be used as marketing instruments for future development projects and that there might be discrepancies between the asked rental and sales prices and the actual transaction prices.[107] With respect to the first of these aspects, i.e. duplicates, the data received from the Real Estate Pilot AG was already processed to exclude duplicates.[108] This was done using a multistage process that considers the type of offer (rent or sale) as well as the information on various characteristics that is available. Overall, it can be said that the type of building, the address, the amount of living space and the price play an important role in this exclusion process. When deciding between two offers for the same building, the one with more information entered is chosen. A blending of the information provided in two different offers does not take place and no other substantial changes are done regarding the micro-data. With respect to the second of theses aspects, i.e. the use as a marketing instrument for future development projects, this could be partially resolved by only including buildings up to a construction year of 2021. The construction year 2021 is included since observations towards the end of 2020 may include newly constructed buildings to be finished in 2021. It seems plausible that rent as well as sales agreements regarding such buildings have already been signed. Another consideration that accounts for the second aspect is that the regression analysis is performed using EPC data. Observations that do not include data on energy performance are discarded. This also helps to eliminate general advertisements of development firms that cannot provide an energy consumption value since the building does not yet exist. However, this elimination also shows a limitation of the sample used for analysis. It cannot be excluded that information regarding buildings with high energy consumption and a bad EPC level is left out on purpose. Whether this is true cannot be proven and it appears to be a general and very basic limitation of an internet-based data extraction strategy – if a person does not enter critical information, it cannot be collected. A general comparison of the mean of the energy consumption of the sample with the overall building stock would also lack validity since the data used is so current that official statistics for this time period regarding the building stock could not be found. Finally, regarding the third aspect, i.e. discrepancies between asked prices and actual prices: Using appraised values and not transaction prices can be an

[101] Cf. Urban and Mayerl (2018, pp. 116 + 135).
[102] Cf. Kholodilin et al. (2017, p. 3224).
[103] Cf. Hyland et al. (2013, p. 945).
[104] Cf. Real Estate Pilot AG (2020).
[105] Cf. Statista Research Department (2020).
[106] Cf. Real Estate Pilot AG (2020). These include e.g. Immobilienscout, Ebay, Augsburger Allgemeine, Donaukurier.
[107] Cf. Kholodilin et al. (2017, p. 3224).
[108] Cf. Real Estate Pilot AG (2020).

accepted way of substitution because only small discrepancies between the two values exist, especially during upward cycles and in big cities.[109] This is assumed to be the case for the data in this paper and supported by the bulwiengesa real estate price index that has shown an increase for the 16. year in a row.[110] Further, by setting higher prices than the market value, time on market increases and often the later transaction value is lower.[111] Thus, the seller has an incentive to price their building at market value and not above.

5.2. Presentation of the descriptive sample statistics & plausibility check

The data received from the Real Estate Pilot AG covers the time period of January 2019 until December 2020 for the Rhein-Main Region in Germany. The Rhein-Main Region was defined by the counties and cities listed on Statista.[112] In total, two datasets were received. One dataset includes the rental market information for the Rhein-Main Region and the other the sales market information. There are 244 277 rental object observations and 123 308 sales object observations in the raw datasets. First, data were analyzed for plausibility. The individual steps of this assessment are shown in Appendix 15, including the R code used. For the benefit of readers of the code, explanations on the steps performed have been added as comments. As a part of this process, observations with missing values were also deleted. The outcome was a significant drop in observations available for the analysis. The final rental market sample includes 44 442 observations. The final sales market sample includes 31 426 observations. This reflects the fact that many rent and sales offers were incomplete and have considerable improvement potential regarding transparent communication of object characteristics to potential tenants or buyers. In the following, the descriptive sample statistics are presented. These refer to the final samples of 44 442 and 31 426 observations for the rent and sales datasets respectively.

The critical variables in the datasets needed to address the hypotheses developed in this thesis are rent and sales prices and energy efficiency. As the German real estate market is analyzed, all prices in this paper are in Euros. The energy efficiency is defined as the amount of kwh needed per square meter per annum by the building considered. Based on this information, the corresponding EPC level of the building is calculated (see Appendix 15). The information regarding the current energy efficiency levels was taken from the exhibit 10 in the GEG.[113] The calculation of the EPC levels on the basis of energy consumption data was done to circumvent the problem of different regulations in Germany. Over the years, the requirements for the EPC levels have tightened and, thus, the actual energy usage of the buildings needs to be converted into the current EPC level to achieve comparability. The distributions of the observations regarding EPC levels for the rent and the sales datasets are shown in Figures 2 and 3. Two statements can be made based on the distributions:

First, most sales and rent observations have an energy performance corresponding to levels D or E. This is also supported by the summary statistics in Table 1, which indicate that mean energy consumption is equal to EPC level D for the rent and EPC level E for the sales dataset. The second interesting observation can be made by comparing Figures 2 and 3, which reveals that the sales market has more offerings compared to the rent market in the very bad performing levels of G and H as well as in the top performing level of A+. One explanation could be that owners are selling off assets with a very bad energy performance e.g. due to higher risk associated with them. The increased amount of assets that are in the top performing category of A+ could indicate that development and refurbishment firms have already realized the importance of energy performance for the current and future market and have adjusted their projects accordingly. Both explanations seem reasonable but are speculative and interesting topics for future research.

Beyond the price and energy variables, building-specific characteristics are also included in the datasets as well as whether a commission fee must be paid for sales objects. Information regarding the location of the object (postal code) and the upload date of the observations are also part of the datasets. The summary statistics regarding most of the variables are presented in Table 1 and 2. Variables not shown in the table include the postal code of the buildings, the time the offer was first posted online and the type of building (e.g. detached house or apartment).

From the summary statistics in Table 1, it can be concluded that the average rental building in the sample costs 885.26 € of cold rent and 1 074.35 € including operational costs. The average construction year is 1981 and the energy consumption of that building is 117.50 kwh/m^2 per annum and thus equal to an energy performance level of D. Overall, the average 2.73 rooms of the building stretch across 79.44m^2. 13% of buildings have not been occupied before and were newly constructed. 30% of all buildings have been refurbished. While 3% of buildings are furnished, only very few are landmarked buildings. 27% of all buildings are equipped with an elevator and 52% have a parking space available. A comparison of the mean and the median of the rent prices indicate a positive skewness of the data. This is supported by maximum values ranging up to 9 800.00 € and 13 310.00 € for cold and warm rent respectively.

The summary statistics of the sales price sample in Table 2 show that compared to the average rental building, the average building up for sale is much bigger with 141.57 m^2 stretching across 4.92 rooms. The construction year of the buildings is the same (1981) and the energy consumption is on average higher (130.30 kwh / m^2 per annum). This corresponds to an energy performance level of E. 10% of all buildings have not been occupied before and were newly con-

[109] Cf. Henger and Voigtländer (2014, p. 15).
[110] Cf. bulwiengesa AG (2021, p. 1).
[111] Cf. Knight (2002, p. 213).
[112] Cf. Statista Research Department (2021).
[113] Cf. GEG, exhibit 10.

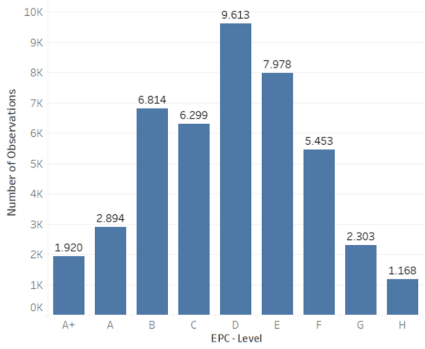


Figure 2: No. of observations in rent sample per EPC – Level (Source: Selfmade)

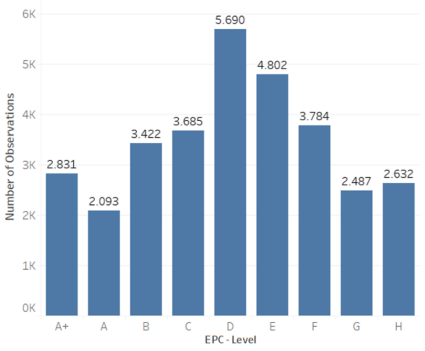


Figure 3: No. of observations in sales sample per EPC – Level (Source: Selfmade)

Table 1: Summary statistics of the rent data sample

Variable	Unit	**Mean**	**Median**	**st.Dn.**	**Minimum**	**Maximum**
Dependent variables						
Cold Rent	Price in Euros (€)	885.26	760 00	493.61	112.80	9 800.00
Warm Rent	Price in Euros (€)	1074.35	940.00	563.49	142.80	13 310.00
Building-specific independent variables						
Energy Consumption	kwh/(m^2 *annum)	117.50	114.10	58.42	5.10	487.40
Living Space	m^2	79.44	74.00	35.63	9.00	707.00
Number of Rooms	Numeric	2.73	3.00	1.11	1.00	12.00
Furnished	Binary, reference = 0; true = 1	0.03	0	0.18	0	1
Refurbished	Binary, reference = 0; true = 1	0.30	0	0.46	0	1
First Occupancy	Binary, reference = 0; true = 1	0.13	0	034	0	1
Landmarked Building	Binary, reference = 0; true = 1	0.00	0	0.01	0	1
Elevator	Binary, reference = 0; true = 1	0.27	0	0.45	0	1
Parking Space	Binary, reference = 0; true = 1	0.52	1	0.50	0	1
Construction Year	Numeric	1981	1984	30.68	1871	2021

Number of observations rent sample: 44 442

structed. 15% of all buildings have been refurbished. This amount of 25% of either newly build or refurbished buildings is much less than the 43% in the rent data. The difference in refurbishment might be a part of the reason why the energy consumption is 12.80 kwh / m^2 per annum lower for the buildings in the rent data. Only 1% of all buildings for sale are a landmarked building. Compared to the rental buildings, a smaller amount of buildings is equipped with an elevator (18%), but more have a parking space available (61%). While 12 % of all buildings are currently being rented by a tenant, 22% of buildings can be purchased without having to pay a commission fee.

Table 2: Summary statistics of the sales data sample

Variable	Unit	**Mean**	**Median**	**st.Dev.**	**Minimum**	**Maximum**
Dependent variable						
Sales Price	Price in Euros (€)	467 315.00	385 000.00	392 455.80	20 000	16 160 000.00
Building-specific independent variables						
Energy Consumption	kwh/(m^2 *annum)	130.30	119.90	80.98	5.30	496.60
Living Area	m^2	141.57	125.00	94.73	13.64	5447.00
Number of Rooms	Numeric	4.92	4.00	3.43	1.00	300.00
Refurnished	Binary, reference = 0; true = 1	0.15	0	0.35	0	1
First Occupancy	Binary, reference = 0; true = 1	0.10	0	0.31	0	1
Landmarked Building	Binary, reference = 0; true = 1	0.01	0	0.08	0	1
Elevator	Binary, reference = 0; true = 1	0.18	0	0.38	0	1
Parking Space	Binary, reference = 0; true = 1	0.61	1	0.49	0	1
Construction Year	Numeric	1981	1983	30.50	1871	2021
Active Lease	Binary, reference = 0; true = 1	0.12	0	0.32	0	1
Contract-specific independent variables						
Commission Free	Binary, reference = 0; true = 1	0.22	0	0.41	0	1

Number of observations sales sample: 31 426

The explanatory variables in the datasets were also checked for correlation. A very high correlation could decrease the significance for both independent explanatory variables considered.[114] The correlation matrix for both samples is shown in Appendix 2 and 3. For the rent data (see Appendix 2), there exists a very high correlation between the living space of a building and the number of rooms (86%). This seems plausible, as both variables are a measure of building size. The question that arises is whether these two variables are valued differently by the market. This might be the case e.g. for city center apartments where a second room might be valued more than having a bigger single room apartment. Further, as the focus of this paper is the impact of energy efficiency, it is more important to control for different characteristics than to minimize correlation between control variables. These two reasons would suggest including both variables in the analysis. A similarly high correlation is present in the sales price data (see Appendix 3) between the number of rooms and the living space (90%). Again, the two reasons mentioned above support including both variables in the analysis. Besides the correlation between these two variables, there is no strong correlation present between explanatory variables. Moderate correlation can be found between the two explanatory variables construction year and energy consumption for the rent data (-55%) and the sales data (-67%). As the energy consumption is the focus of this analysis, this could be problematic regarding the significance of the coefficients. The values computed were also used as a first indicator for multicollinearity (a high correlation coefficient indicates potential multicollinearity). The multicollinearity assumption will be tested in detail after the model specification is defined (see subsection 7.2).

6. Specification of the hedonic price models used for analysis

Based on the theoretical fundamentals of hedonic models explained above, empirical models that can help test the major hypotheses of this paper will now be defined. The heterogeneous good considered is a residential building with certain characteristics and a sales price or rent price. It is essential to identify these characteristics and account for them in the hedonic model. The main categories to consider are a) building-specific characteristics b) location-specific characteristics, c) time-specific characteristics and d) contract-specific characteristics.[115]

The three hypotheses of interest need to be investigated using three different models. The reason for this is that each hypothesis looks at the impact of energy efficiency on the valuation of a building on a different level. Each level has a different dependent variable. In hypothesis A, the net operating income is considered. This will be tested using the cold rent of a building as the dependent variable in the hedonic model. An increase in cold rent, c.p., is equivalent to an increase in the NOI of the building. Hypothesis B considers the GEI of the building. Here, the dependent variable will be the warm rent of a building. As the warm rent is the total incoming cash flow, an increase is, c.p., equivalent to an increase in GEI. The dependent variable of the hedonic model for hypothesis C will be the sales price of a building. Based on the DCF method discussed above, an implication of hypotheses A and B is that the value and therefore the sales price of the building is, c.p., increased. This implication would have to be visible in the sales price dataset. To make the following empirical analysis as well as the discussion more intuitive, the models will be named "Cold Rent Model", "Warm Rent Model" and "Sales Price Model". All models include data transformations to improve interpretation and compliance with the regression assumptions as well as dummy control variables to account for location and time effects. Of note, the indexes used in the equations and the meaning of the variables remain the same for all three models. Details on the coding and data transformations and an overview of which variables are used in each model are summarized in Appendix 1.

The Cold Rent hedonic model is expressed by the following equation:

$$\begin{aligned}\ln\left(cold\_rent_{ilt}\right) &= \alpha + \beta_1 epc\_level_i \\ &+ \beta_2 \ln\left(living\_space_j\right) + \beta_3 no\_rooms_i + \beta_4 furnished_i \\ &+ \beta_5 refurbished_i + \beta_6 first\_occupancy_i \\ &+ \beta_7 landmarked\_building_i + \beta_8 elevator_i \\ &+ \beta_g parking\_space_i + \beta_{10} building\_type_i \\ &+ \beta_{11} construction\_year_i + \gamma_l + \delta_t + \varepsilon_{ilt}\end{aligned} \tag{5}$$

The dependent variable, *"cold_rent"*, is transformed as the natural log. This helps to account for potential non-linearity present in the model.[116] The implications for the interpretation of the results will be discussed later. They also depend on the independent variable considered and whether it is transformed or not. In the model, i is the index for the single building observation. The index l defines the location while the t index defines the date of the observation. The constant in the model is represented by α. This is followed by the various building-specific characteristics that were already mentioned in the summary statistics of the data above, a control variable for the building type (*"building_type"*), the location effects ("γ_l"), the time effects ("δ_t") and the error term ("ε_{ilt}"). The independent variable relevant for the assessment of the hypothesis is the first building-specific characteristic, i.e. a dummy variable indicating the EPC-level of the building. The EPC level of the building is equivalent to the one calculated on the basis of the energy consumption of the building (see subsection 5.2). The reference level of this variable is set to D. Any comments regarding energy

[114] Cf. Fahrmeir, Kneib, and Lang (2009, p. 154).
[115] Cf. Sopranzetti (2010, pp. 1202–1203).
[116] Cf. Fahrmeir et al. (2009, pp. 22-23).

(in)efficiency are made with respect to this reference level. This was also recommended by others to increase comparability between studies.[117] The quantitative variable *"living_space"* in m^2 is log-transformed. This is again done to account for non-linearity.[118] The other building specific characteristics are all categorical as well as dummy variables. The *"no_rooms"* variable is coded as a categorical variable to account for premiums or discounts associated with different levels of this variable. The variables *"furnished", "refurbished", "first_occupancy", "landmarked_building", "elevator" and "parking_space"* only have two levels, zero and one. Zero is the reference value. This would mean that the respective building characteristic is not applicable for this observation. The variable *"building_type"* has ten different levels and serves as a control variable for different building types, e.g. the difference between a detached house and an apartment. The last building-specific variable, i.e. *"construction_year"*, helps to control for depreciation of the building. Following Cajias et al. (2019), the construction dummies are coded in ten-year steps.[119] The remaining two variables in the Cold Rent Model account for the location effects of the building on a postal code level ("γ_l") and the time effects on a monthly level ("δ_t").

The equation of the warm rent model uses *"warm_rent"* as dependent variable. This results in the following equation:

$$\begin{aligned}\ln\left(warm\_rent_{ilt}\right) &= \alpha + \beta_1 epc\_level_i \\ &+ \beta_2(\ln living\_space_i) + \beta_3 no\_rooms_i + \beta_4 furnished_i \\ &+ \beta_5 refurbished_i + \beta_6 first\_occupancy_i \\ &+ \beta_7 landmarked\_building_i + \beta_8 elevator_i \\ &+ \beta_g parking\_space_i + \beta_{10} building\_type_i \\ &+ \beta_{11} construction\_year_i + \gamma_l + \delta_t + \varepsilon_{ilt}\end{aligned} \tag{6}$$

The empirical analysis in this paper starts with NOI and then looks at the GEI before considering the sales price. Since a comparison between the different results is needed, it would not be meaningful to generate a fundamentally different model for the analysis of the sales price. However, it seems expedient to include or exclude variables that are only relevant for the respective transaction (rent or sale). Therefore, a variable indicating the current rent status (*"rent_status"*) of the building is added to the Sales Price Model as well as a control variable for the contract-specific characteristic regarding the sales commission (*"commission_free"*). The variable *"rent_status"* is a two-level dummy variable with a reference value of zero indicating that no one is currently renting the building. The contract-specific variable is a two-level dummy variable with zero, not commission free, as the reference value. The variable *"furnished"* is dropped, as this is assumed to be an uncommon feature for buildings that are sold and not let. This leads to the following equation:

$$\begin{aligned}\ln\left(sales\_price_{ilt}\right) &= \alpha + \beta_1 epc\_level_i \\ &+ \beta_2 \ln\left(living\_space_i\right) + \beta_3 no\_rooms_i \\ &+ \beta_4 rent\_status_i + \beta_5 refurbished_i \\ &+ \beta_6 first\_occupancy_i + \beta_7 landmarked\_building_i \\ &+ \beta_8 elevator_i + \beta_g parking\_space_i \\ &+ \beta_{10} building\_type_i + \beta_{11} construction\_year\_ \\ &+ \beta_{12} commission\_free_i + \gamma_l + \delta_t + \varepsilon_{ilt}\end{aligned} \tag{7}$$

7. Presentation of the empirical results & assessment of the model assumptions

7.1. Empirical results of the hedonic regression models

As the level of analysis differs for each model, the empirical results should also show different effects of energy efficiency on the rent or sales price. If the hypotheses tested in this paper are true, a positive impact of energy efficiency should be visible and significant in all models. Further, the effect should be rather small for the Warm Rent Model and somewhat bigger for the Cold Rent Model because of the capitalization of energy savings. For the Sales Price Model, the effect should be the largest, as additional factors such as risk of depreciation are included. Energy inefficiency on the other hand should have a negative impact on the rent and sales prices. Here, cold rent should decrease to account for higher operational costs while staying competitive with the warm rent in the market. The question arises whether signaling effects are also present for these buildings. With negative signaling effects being present for energy inefficient buildings, the warm rent of these buildings should be lower compared to buildings of reference level D. The sales price of an energy inefficient building should also decrease. The magnitude of the decrease of the sales price should, similar to the efficiency premium, be bigger than the decrease in cold rent as additional factors such as exposure risk to future regulation changes are increased. In the remainder of this subsection, the empirical results of the models are presented. After the presentation of results, the assumptions regarding the linear models are discussed.

Table 3 shows the results of the linear models defined by equations (5) – (7). All the dependent variables were transformed using the natural logarithm and, thus, the coefficient of the estimator of an explanatory variable is equal to the increase of the natural log of the dependent variable. To facilitate the understanding of the economic meaning of the results, coefficients are converted into percentage values in the text. The non-converted values can be found in the respective table for comparison purposes. The standard errors reported in the table are beneath the coefficients and robust White standard errors that correct for heteroscedasticity.[120] See Appendix 15 for computational details.

[117] Cf. Cespedes-Lopez et al. (2019, p. 53).
[118] Cf. Fahrmeir et al. (2009, pp. 22-23).
[119] Cf. Cajias et al. (2019, p. 184).
[120] Cf. Fahrmeir et al. (2009, pp. 135 – 136).

The Cold Rent Model shows an overall statistical significance between the independent explanatory variables and the dependent variable cold rent (F (539, 43 902) = 776.20, p < 2.2e-16, R^2 = .9050). The F-statistic is highly significant with a value of 776.20 and there is only a chance of less than 2.2e-16 that the Cold Rent Model does not have any explanatory power.

The degrees of freedom equal 539 for the regression and 43 902 for the error. Further, 90.50% of the variance present in the data can be explained by the model. As the model includes various explanatory variables, the adjusted R^2 is also considered. This is done to account for the possibility that a high number of explanatory variables is the cause for the high R^2 value.[121] The Cold Rent Model shows an adjusted R^2 of 0.9039. As this value is not very different from the R^2 value (R^2 = .9050), it is highly unlikely that the model includes variables that increase the value of the explained variance only by chance. When looking at the independent explanatory variables, a significant cold rent price premium is present for EPC levels above reference level D. For EPC levels A+, A, B and C the cold rent premium is equal to 5.82%, 2.04%, 3.06% and 0.69% respectively. For A+, A and B this finding is highly significant at the 0.1% level. For EPC level C this finding is significant at the 1% level. The magnitudes of the coefficients below the EPC level of D are much smaller, but all of them are negative. The overall discounts are -0.58%, -0.28%, -0.70% and -0.15% for the EPC levels of E, F, G and H respectively. The values below D do not show a clear linear decrease of cold rent and indicate that the magnitude of the effect of energy efficiency on cold rent might be smaller compared to EPC values above D. The significance of results is smaller with only E being significant at a 5% level. Thus, it cannot be excluded that EPC values of F, G and H have no impact on the cold rent of the building. Overall, the Cold Rent Model shows a significant premium for energy efficient homes of up to 5.82% but does not indicate significant cold rent discounts for energy inefficient buildings.

The Warm Rent Model shows an overall statistical significance between the independent explanatory variables and the dependent variable warm rent (F (539, 43 902) = 796.20, p < 2.2e-16, R^2 = .9072). The F-statistic is highly significant with a value of 796.20 and there is only a chance of less than 2.2e-16 that the Warm Rent Model does not have any explanatory power. The degrees of freedom equal 539 for the regression and 43 902 for the error. Further, 90.72% of the variance present in the data can be explained by the model. As the model includes various explanatory variables, the adjusted R^2 is also considered. The Warm Rent Model shows an adjusted R^2 of 0.9060. As this value is not very different from the R^2 value (R^2 = .9072), it is highly unlikely that the model includes variables that increase the value of the explained variance only by chance. When looking at the independent explanatory variables, a significant warm rent price premium is present for some of the EPC levels that are above the reference level D. For EPC levels A+, A, B and C the coefficient of the estimator is equivalent to a 3.86%, 0.38%, 1.98% and 0.21% increase in warm rent respectively. For A+ and B this finding is highly significant at the 0.1% level. The magnitudes of the coefficients below the EPC level of D are all smaller than 1% and not significant. The coefficients do not show a clear linear trend. Compared to the coefficients of the Cold Rent Model, they are smaller and closer to zero. An effect of lower EPC levels on warm rent cannot be assumed. Overall, the Warm Rent Model shows a significant premium for energy efficient homes in the categories A+ and B with a maximum of 3.86% and no significant discounts for the energy inefficient buildings below the reference level of D.

The Sales Price Model shows an overall statistical significance between the independent explanatory variables and the dependent variable sales price (F (605, 30 820) = 336.70, p < 2.2e-16, R^2 = .8686). The F-statistic is highly significant with a value of 336.70 and there is only a chance of less than 2.2e-16 that the Sales Price Model does not have any explanatory power. The degrees of freedom equal 605 for the regression and 30 820 for the error. Further, 86.86% of the variance present in the data can be explained by the model. As the model includes various explanatory variables, the adjusted R^2 is also considered. The Sales Price Model shows an adjusted R^2 of 0.8660. As this value is not very different from the R^2 value (R^2 = .8686), it is highly unlikely that the model includes variables that increase the value of the explained variance only by chance. When looking at the independent explanatory variables, a significant sales price premium is present for some of the EPC levels that are above the reference level of D. For the EPC levels of A+, A and B the sales premium is significant and equal to an increase of 6.81%, 3.14% and 1.52% of the sales price respectively. For A+ and A this finding is highly significant at the 0.1% level and for B at the 5% level. The estimated coefficient for the EPC level C is equal to 0.09% and not significant. When considering the EPC levels below the reference level of D, the coefficients are equal to -0.70%, 0.69%, -1.73% and -8.80% for E, F, G and H respectively. The finding for EPC level G is significant at the 5% level and the finding for the EPC level H is significant at the 0.1% level. For EPC levels of C, E and F no significant difference was seen. Thus, these EPC levels do not seem to have an impact on the sales price of a building. Overall, the Sales Price Model shows a significant premium of up to 6.81% for energy efficient buildings with an EPC level of B and above. It finds neither a significant discount nor premium for buildings with an EPC level ranging from C to F. Starting with G, the Sales Price Model finds significant discounts for energy inefficient buildings with discounts of up to -8.80% for EPC level H.

The empirical results regarding the control variables of the models are also highly significant. In the Cold Rent Model, there is a significant increase in rent for more living space (100% increase in living space, increases cold rent by 73.19%) and for furnished apartments (20.27% increase). Renting out newly constructed buildings comes with a premium of 8.92%, while refurbished apartments are 4.00%

[121] Cf. Urban and Mayerl (2018, p. 91).

Table 3: Empirical results of hedonic regression models

Dep. Var.:	**Cold Rent**	**Warm Rent**	**Sales Price**
	(1)	(2)	(3)
EPC Level A+	0.05660055***	0.037870311***	0.065838973***
	0.00437359	0.004078778	0.007952356
EPC Level A	0.02021563***	0.003763821	0.030892444***
	0.00396247	0.003716371	0.008115774
EPC Level B	0.03010108***	0.019579393***	0.015110583*
	0.00306967	0.002882431	0.006249028
EPC Level C	0.00682597***	0.002068312	0.000932055
	0.00247206	0.002320161	0.004804528
EPC Level E	−0.00583240*	−0.001800472	−0.007051507
	0.00227909	0.002128358	0.004499839
EPC Level F	−0.00277893	0.004343243	0.006825412
	0.00267550	0.002493251	0.005271737
EPC Level G	−0.00702567	0.000472069	−0.017478196*
	0.00382009	0.003588644	0.006900535
EPC Level H	−0.00145708	−0.001764933	−0.092088491***
	0.00541184	0.005129463	0.007701151
Ln(Living Space)	0.79237579***	0.762227770***	0.860093156***
	0.00459536	0.004387214	0.008272612
Furnished	0.18452401***	0.174828571***	-
	0.00558676	0.005187815	
Refurnished	0.03925094***	0.034063595***	0.051384371***
	0.00168391	0.001576583	0.003985229
First Occupancy	0.08546231***	0.074777163***	0.048026681***
	0.00234633	0.002189902	0.005424765
Landmarked Building	0.01843539	0.038014154	0.047517962***
	0.02055397	0.028272040	0.019778178
Elevator	0.02467841***	0.040704650***	−0.016977286***
	0.00201453	0.001862656	0.004077302
Parking Space	0.03101226***	0.030620634***	0.007289280*
	0.00165583	0.001546589	0.003011203
Rent Status	-	-	−0.060121281***
			0.004287139
Commission Free	-	-	0.015091733***
			0.003507405
Intercept	2.64195778***	2.910135919***	7.929795028***
	0.09239100	0.115792996	0.053315457
Categorical Control Variables			
No. of Rooms	1	1	1
Building Type	1	1	1
Construction Year	1	1	1
Location	1	1	1
Upload Date	1	1	1
R squared	0.9050	0.9072	0.8686
Adjusted R squared	0.9039	0.9061	0.8660
No. of observations	44 442	44 442	31 426

Significance level: (*) $p < 0.05$; (**) $p < 0.01$; (***) $p < 0.001$

more expensive than non-refurbished ones. A landmarked building does not provide any significant value, while an elevator increases cold rent by 2.50% and an available parking space increases cold rent by 3.15%. In the Warm Rent Model, the magnitude of significant coefficients decreases for all control variables except for the elevator. An elevator being present increases warm rent by 4.15% compared to the 2.50% in the Cold Rent Model. This seems plausible, as an elevator has an impact on the operating costs, and thus its relative influence on the rent increases when considering the warm rent. In the Sales Price Model, the control variables are also highly significant. A 100% increase in living space increases the sales price by 81.52%. A refurbished building is valued 5.27% higher by the market while a newly built building has a 4.92% higher sales price. Regarding the sales price, landmarked buildings have a premium of 4.87%. This might be the case because they may have a historic importance that is valued with a premium. An elevator decreases the sales price of a building by -1.68 % while an available parking space only comes with a premium of 0.73%. A building with an active lease agreement costs 6.20% less and a commission free building is offered for 1.52% more on the market.

7.2. Assessment of the model assumptions

To assess the explanatory power of the three different models, the assumptions underlying multiple regression need to be validated. Each assumption will now be considered for all three models.

1. Linearity: The estimated coefficients are of a linear nature.

The linearity assumption is tested by plotting the residuals of the models on the y-axis against the fitted values on the x-axis.[122] Doing this can help detect previously overlooked non-linear influences of explanatory variables. The plot of the Cold Rent Model (see Appendix 4) shows randomly distributed residuals for the most part. When looking at the highest as well as lowest fitted values, there is a small indication of a weak quadratic relationship also shown by the fitted line. As this deviation from linearity is insignificantly small, the linearity assumption is considered fulfilled. Like the plot of the Cold Rent Model, the plot of the Warm Rent Model (see Appendix 5) shows a small deviation at the top and bottom of fitted values. Still, the assumption of linearity for the Warm Rent Model is approximately fulfilled. The plot of the Sales Price Model (see Appendix 6) shows randomly distributed residuals and an almost perfect horizontal line indicating a linear relationship. The linearity assumption is fulfilled for the Sales Price Model.

2. Exogeneity: The mean of the error term is equal to zero.

According to Urban and Mayerl (2018) the exogeneity assumption cannot be tested as the expected value of zero for the mean of the error term refers to the actual population model.[123] This model is unknown. Only an estimation of the population model based on a sample, the regression model, is known. For technical reasons, the mean of the error term is always equal to zero for the regression model. Not having a mean equal to zero of the error term would only bias the intercept estimation.[124] As the focus of this thesis is the impact of energy efficiency, a potentially biased intercept estimation would not impact the results regarding the EPC levels. Thus, even in this worst case, the findings of this thesis would still be valid.

3. Homoscedasticity: The variance of the error term is constant.

Not fulfilling the homoscedasticity assumption has an impact on the validity of the significance tests of the linear model. Thus, with strong heteroscedasticity present in the model, no statement can be made regarding the significance of estimated coefficients.[125] To minimize this effect, the dependent variables are log-transformed.[126] Again, residual plots can be used for diagnosis (see Appendix 7 - 9).[127] Here, the square root of the absolute values of the standardized residuals are plotted on the y-axis and the fitted values of the models are plotted on the x-axis. A random and linear distribution around the value of 1 is indicative of homoscedasticity. All three models show a slight deviation from linearity. This is a known limitation of multiple linear regression and it is recommended to take a heuristic approach for diagnosis by considering statistical test values (e.g. Breusch Pagan test) and graphical representations.[128] Nonetheless, to ensure that the significance of the test statistics regarding the linear models can be regarded as valid, robust White standard errors were computed and are presented in the Table 3 with the empirical results. The use of this method is recommended as it also does not need information regarding the form of heteroscedasticity.[129] Based on heuristic arguments and the additional computation and use of robust White standard errors, the homoscedasticity assumption for all three models can be seen as sufficiently fulfilled or, at the very least, not critical for the interpretation of the results of the model.

4. Autocorrelation: There exists no covariance between error terms.

The Durbin-Watson test is often used for diagnosing autocorrelation.[130] The value of the test statistic ranges from

[122] Cf. Urban and Mayerl (2018, pp. 199-200).

[123] Cf. Urban and Mayerl (2018, p. 196), also for the following 3 sentences.

[124] Cf. Urban and Mayerl (2018, p. 196).

[125] Cf. Urban and Mayerl (2018, pp. 253-254), also for the previous sentence.

[126] Cf. Fahrmeir et al. (2009, p. 132).

[127] Cf. Urban and Mayerl (2018, p. 299).

[128] Cf. Fahrmeir et al. (2009, pp. 131–132).

[129] Cf. Fahrmeir et al. (2009, pp. 135 – 136).

[130] Cf. Fahrmeir et al. (2009, pp. 141 – 142).

0 to 4 with the middle value of 2 indicating no autocorrelation. Further, the null hypothesis states that autocorrelation is equal to zero while the alternative hypothesis states that autocorrelation is unequal to zero. The computation is equal to 1.98 and significant at the 5% level ($p = 0.012$) for the Cold Rent Model, 1.98 and significant at the 5% level ($p = 0.046$) for the Warm Rent Model and 1.99 and not significant ($p = 0.422$) for the Sales Price Model. The computation was done using the "durbinWatsonTest" function from the "car" package in R. The result for the Sales Price Model is clear: No autocorrelation present and a value of close to 2. The test for the Cold and Warm Rent Models also show values very close to 2. However, here the significance at the 5% level indicates that the null hypothesis has to be rejected. Because of the test value being very close to 2, this result needs to be interpreted. In case of large datasets, statistical tests may become significant even with very small effect sizes.[131] This is the case here and the test has detected a minor autocorrelation. Although this may be present, the fact that the values are almost ideal, i.e. close to 2, shows that this autocorrelation will not have a major impact on the interpretation of the results and can be discounted accordingly.

5. Multicollinearity: There exists no perfect multicollinearity between explanatory variables.

As a first indication, the correlation matrix was used. Several correlations between variables were found (see Appendix 2 and 3) indicating potential multicollinearity. Multicollinearity can impact the stability of the estimated coefficients.[132] Thus this needs to be investigated further. It is suggested taking the variance inflation factor (VIF) into account when diagnosing multicollinearity.[133] However, the models used in this analysis include various categorical variables with more than one degree of freedom. This leads to a measure of collinearity that is partly artifactual.[134] To overcome this problem, the generalized VIF (GVIF) and $GVIF^{(1/(2*DF))}$ for the three linear models were computed. The values of the $GVIF^{(1/(2*DF))}$ were used, as is suggested.[135] Fahrmeir et al. (2009) state that for any VIF higher than 10 there is a significant multicollinearity problem.[136] No value indicating problematic multicollinearity was found (see Appendix 13-14). In particular, the values for construction year and energy efficiency were found to be small. The computations were done using the "vif" function of the "car" package in R. Based on these calculations, the multicollinearity assumption is sufficiently fulfilled for all three models.

6. Normality of residuals: The error terms are normally distributed.

The normal distribution of residuals is important for the explanatory power of the test statistics.[137] A graphical analysis of this assumption can be done using the quantile-quantile plot.[138] Here, the standardized residuals are plotted on the y-axis while the theoretical quantiles are plotted on the x-axis. Based on this graph, all three models show good normality for most of the data (see Appendix 10 - 12). At the bottom and top theoretical quantiles, the normality assumption is violated. However, in the case of large samples, conclusions drawn from the models are still valid: Fahrmeir et al. (2009) argue that with a large number of observations, the OLS estimator approximately shows the same normal distribution as with fulfilling the normality assumption of the error term.[139] This is important for the validity of the test statistics. The needed conditions for this approximation can be seen as fulfilled if the observations and thus their variables come from a random sample, as is the case in this empirical analysis.[140] The examples used by Fahrmeir et al. (2009) include samples with less than 5000 observations.[141] With six-fold larger sample sizes in the present analysis, the approximation is possible. In conclusion, even though the normality assumption of the error term is not fulfilled, we can assume a normal distribution for the OLS estimator.

8. Discussion of the empirical results & limitations of the thesis

In the following, the empirical results of the three hedonic models are discussed in the context of the literature. The limitations of this thesis work are also considered.

Hypothesis A states that the net operating income (NOI) is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building and the valuation of the building is increased or decreased respectively. To analyze the first part of this hypothesis, the Cold Rent Model was defined. Assuming that the participants of the rental market act purely rational, there should be a clearly visible positive linear trend from EPC level H to A+ as a landlord will recoup energy cost savings by increasing cold rent. For EPC levels above the reference value D, this hypothesis is supported by the empirical results (see Table 3). For EPC values E-H there is no clear relationship between EPC levels and cold rent, suggesting that landlords of H rated buildings can charge the same cold rent as landlords of D rated buildings despite of higher energy costs. Since it is unlikely that rational acting market participants are indifferent to higher energy costs, this finding may reflect the tight real estate market in the Rhein-Main Region[142] and, thus, tenants, must accept higher operational costs. Furthermore, with no clear discount present for energy inefficient houses, a landlord has little incentive to improve the EPC level of e.g. a H rated building to a D rated building. For a landlord to see financial benefit, EPC ratings of the upper segments A+ through

[131] Cf. Urban and Mayerl (2018, p. 131).
[132] Cf. Urban and Mayerl (2018, p. 252).
[133] Cf. Fahrmeir et al. (2009, pp. 170 – 171).
[134] Cf. Fox and Monette (1992, p. 180).
[135] Cf. Fox and Monette (1992, p. 180).
[136] Cf. Fahrmeir et al. (2009, p. 171).
[137] Cf. Urban and Mayerl (2018, p. 187).
[138] Cf. Fahrmeir et al. (2009, p. 169).
[139] Cf. Fahrmeir et al. (2009, pp. 105 – 106).
[140] Cf. Fahrmeir et al. (2009, pp. 105 – 106).
[141] Cf. Fahrmeir et al. (2009, p. 5).
[142] Cf. Manus (2020). Cajias et al. (2019, pp. 186-187).

C need to be achieved. Whether such investments are financially viable needs to be assessed in future research. Thus, hypothesis A is partially supported by data: better energy efficiency c.p. increases cold rent and NOI for buildings rated above D. Less energy efficient buildings show no significant discount. The implications of these findings for building valuation are straightforward: Above D rated buildings should be valued higher because of a higher NOI and below D rated buildings should receive no or only small discounts regarding their valuation. This will be discussed further in the context of hypothesis C.

Hypothesis B states that the gross effective income (GEI) is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building. To analyze this hypothesis, the Warm Rent Model was used and its findings were compared with those of the Cold Rent Model (previous paragraph). Since warm rent is equal to cold rent plus all operational costs the tenant is charged, and since c.p. operational costs should rise with energy inefficiency, i.e. lower EPC levels, the Warm Rent Model combined with the Cold Rent Model provides some insight into the allocation of operational costs by landlords. As far as energy efficient buildings are concerned, the Warm Rent Model revealed significant warm rent premiums for EPC levels of A+ and B, indicating that c.p. the increase in cold rent exceeded the reductions in operating costs of the building. This indicates that additional factors are at play enabling the landlord to charge a premium, e.g. signaling effects. Although the present analysis does not identify these factors, it is clear that they add value for the tenants as the D and A+ rated buildings compete in the same market. Of note, this effect is only present for EPC levels of A+ and B. EPC levels A and C were not significantly different from D, indicating that for such buildings the landlord recouped the energy cost savings by increasing the cold rent (see Table 3 and compare (1) and (2)). At the other side of the scale, i.e. EPC levels below D, no coefficients are significant. This is a surprising finding because the increasing energy consumption from D to H neither affected cold nor warm rent. The reasons for this are unclear and a more detailed study taking the exact mix of operating costs of the different buildings into account will be needed. Some follow up hypotheses that could be addressed to better understand the absence of an increasing warm rent at low rated EPC building levels are:

- the landlord financially offsets some of the operating costs for energy inefficient buildings so that the building stays competitive in the rental market
- the operating costs entered in the rental offer is lower than the actual operating costs charged by the landlord
- the actual energy consumption and operating costs of energy inefficient buildings is not higher than D rated buildings because higher energy costs are offset by a reduction in other operating costs (e.g. lower technology level with no IT infrastructure)

Hypothesis C states that the market value is – c.p. – higher for a more energy efficient and lower for a less energy efficient residential building and the increase/decrease is proportionally bigger than the increase/decrease in NOI. To test this hypothesis, the Sales Price Model was defined. The empirical results (see Table 3) show that there is – as predicted – a premium for more energy efficient buildings (see EPC levels A+, A, B) and a discount for energy inefficient buildings (see EPC levels G and H). However, these effects are only seen at each end of the scale. For a wide range of energy consumption values, the sales price does not show a premium or discount (see EPC levels C – F). Of note, the empirical results of the Sales Price Model reflect the results of the rent models only at the energy efficient end of the scale. Energy inefficient buildings, which did not show a discount for rent, showed a significant discount for sales prices. This suggests that other factors, e.g. asset risk factors could be involved. It appears plausible that potential buyers see an increased asset risk[143] in the energy inefficient buildings, since future policy changes may force them to invest heavily into their property to make them more energy efficient. In the mid-range other factors could play a role besides investment principles since the owner-occupier rate is higher than 45% in Germany[144] and they may prioritize other building aspects over its energy efficiency.

As with all empirical studies, there are limitations. First, the sample used for this analysis only includes data from the Rhein-Main Region. More regional analyses need to be performed to understand whether differences are present between regional markets. This will help to assess the generalizability of the results. Second, the sample consists of rent and sales offers. This limitation was already discussed in subsection 5.1 regarding the data generating process. Data from real transactions should yield results with greater generalizability. Such data are currently not publicly available. Third, many observations were lost because offerings were incomplete and missing values regarding the energy consumption or other variables of the buildings led to their subsequent deletion from the data set used for analysis.

In line with previously published literature, this thesis supports the conclusion that energy efficient residential buildings are sold and let for a premium.[145] It now shows this for the Rhein-Main Region in Germany using the current energy efficiency classification for buildings. Compared to Cajias and Piazolo (2013), who also based their analysis on data from Germany, this thesis shows a smaller impact of energy efficiency on the rent and sales prices.[146] Further, this study shows – at least for the Rhein-Main Region in Germany – that the impact of the categorical variable *"epc_level"* cannot be described as linear, as suggested by other authors.[147]

[143] This would entail an increased discount rate and subsequently a decrease in valuation of the building.
[144] Cf. Statistisches Bundesamt (2020).
[145] Cf. e.g. Cajias and Piazolo (2013, p. 53). Cajias et al. (2019, p. 189).
[146] Cf. Cajias and Piazolo (2013, p. 65).
[147] Cf. Kholodilin et al. (2017, p. 3231).

In this earlier study, an increase of 1 kwh/m^2 per annum was calculated to decrease the sales price by 0.05% and rent by 0.02%.[148] Comparing a D rated building with an A+ rated building (difference of around 100 kwh/m∧2 per annum), this would be equal to a 2% rent price and a 5% sales price premium. The findings in this thesis show values that are larger for the rent and the sales price premium. This difference between the analysis presented here and the earlier analysis could be the result of the linear and categorical descriptions or be caused by an increased awareness regarding climate and environment since the study was performed. Indeed, a few years later, but still before the most recent awareness shift, small effects of energy efficiency on the rental prices in regions such as Frankfurt were shown,[149] speaking for the latter hypothesis. Compared to the findings of this second study, the current analysis shows stronger impacts of energy efficiency, especially for the EPC levels of A+ through B. This may be the result of the level of analysis since nationwide data were analyzed[150] or it may be the result of a real change during the last few years driven by the change in climate awareness.

Unfortunately, there are nation-specific implementations of the EPC which make it difficult to directly compare the results of this study with results from other European countries. A comparison with studies not using the EPC as a proxy is even more limited. With all necessary caution, it can be said the sales prices in the Rhein-Main Region show a smaller premium compared to the sales price premium reported for other European countries. For example, the 6.5% premium found here is smaller than the 10.2% for the best performing EPC level found in the Netherlands[151] and the 9% premium found in Ireland.[152] Thus, the effect of energy efficiency on building valuation shows clear nation- and region-specific differences, which need to be considered by actors in these specific markets. It would be interesting to follow these nation-specific developments to find out if European policy changes will harmonize these developments across the European real-estate market.

9. Conclusion & outlook

The fundamental questions addressed by this paper are (1) whether and (2) how energy efficiency changes building valuation in the residential sector. To assess the potential impact of energy efficiency on building valuation, three different target variables were identified and used as read outs: The NOI, the GEI, and the market value. These target variables were chosen based on the valuation fundamentals of real estate. If energy efficiency has an effect on these target variables, it will be indicative that the energy efficiency of buildings influences their valuation. Further, the sign of the target variables, i.e. positive or negative, will show in which direction energy efficiency affects the valuation. Based on these deliberations, testable hypotheses were formulated for the NOI, GEI, and market value respectively (see p. 10). To address these hypotheses, three hedonic regression models were generated and two large empirical datasets of real estate offerings in the Rhein-Main Region in Germany were analyzed. The empirical results were presented and discussed in the context of the literature. In the following, this thesis concludes by summarizing the main implications of the findings for the real estate market and policy makers and the potential for future research.

This paper contributes to the literature with a detailed analysis of how the energy efficiency of buildings impacts their valuation in the residential real estate market of the Rhein-Main Region, one of the metropolitan areas of Germany, using the most recent data available. The study captures the impact of recent trends in the finance industry and in the German society in general on the real estate market of this region, making the results of the study relevant for stakeholders in the residential real estate market, e.g. developers, investors and regulators. Further, this paper has identified several areas of research that are of interest to understand the difference in or the non-existence of premiums/discounts in the future. In a nutshell, the following conclusions can be drawn:

- Landlords of average buildings can improve their current rental income by investing in energy efficiency. This investment shows a significant additional return if an A+ or B EPC level can be achieved. This premium goes beyond recouping energy cost savings and needs to be included when checking investments for financial viability.
- Developers can increase sales prices of their buildings by increasing energy efficiency above an EPC level of C. This needs to be considered in the profitability analysis. If the additional cost of construction for achieving EPC level A+, A or B is lower than the premium achieved, profitability can be increased.
- Private and institutional owners of G and H rated buildings should consider improving the energy efficiency of these buildings to increase their value and reduce exposure to future risks.
- Regulators and policy makers in Germany need to improve data transparency, availability and consistency regarding the residential real estate market to make financial implications of energy (in)efficiencies more visible.
- Regulators and policy makers on a European level need to introduce and implement a homogeneous EPC rating to increase comparability between literature and the different real estate markets in general.

[148] Cf. Kholodilin et al. (2017, p. 3231).
[149] Cf. Cajias et al. (2019, p. 187).
[150] Cf. Cajias et al. (2019, p. 182).
[151] Cf. Brounen and Kok (2011, p. 176).
[152] Cf. Hyland et al. (2013, p. 950).

Future research in this area should analyze the impact of energy efficiency in greater detail and should take the specific differences present in the real estate market and in societies into account. Different perspectives should be explored that will help to refine the results. These include but are not limited to looking at differences regarding energy efficiency premiums for different building types, the impact of socioeconomic factors on these premiums, differences between the energy source used in the building (e.g. oil, gas or pellets etc.) and the impact of CO_2 prices on energy efficiency premiums.

The underlying theme of sustainability will, based on current developments, very likely become more important in the years to come. As a direct consequence of this societal development, the energy efficiency of buildings will also become more relevant and differences in valuation will be even more pronounced in the future. The change in the real estate market has just begun as much of the current building stock has to be refurbished to reach climate targets set by the EU.[153] This also makes stricter regulations for new builds and a rise in cost for CO_2 intensive energy sources likely. Research in this area is urgently needed and this thesis may have contributed to this discussion by highlighting the effects of the energy efficiency of buildings on their valuation in the residential real estate market in a major metropolitan region of Germany.

[153] Cf. European Commission (2020a, p. 1).

References

Ankamah-Yeboah, I., Rehdanz, K., et al. (2014). *Explaining the variation in the value of building energy efficiency certificates: A quantitative meta-analysis* (Tech. Rep.). Kiel Working Paper.

Baum, A. E., & Hartzell, D. (2021). *Real estate investment: Strategies, structures, decisions*. second edition, Chichester, Wiley Online Library.

Belke, A., & Keil, J. (2017). Fundamental determinants of real estate prices: A panel study of german regions. *Ruhr Economic Papers*, *731*, 1–29.

Brounen, D., & Kok, N. (2011). On the economics of energy labels in the housing market. *Journal of environmental economics and management*, *62*(2), 166–179.

Brown, M. J., & Watkins, T. (2016). The "green premium" for environmentally certified homes: a meta-analysis and exploration. *Working Paper, Oregon Department of Environmental Quality*.

bulwiengesa AG. (2021). *Marktstudie bulwiengesa – immobilienindex 1975 – 2020.* Retrieved from https://www.bulwiengesa.de/sites/default/files/immobilienindex_2021_0.pdf (Retrieved 18.02.2021)

Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz, & Bundesumweltamt. (2020). *Umweltbewusstsein und umweltverhalten.* Retrieved from https://www.umweltbundesamt.de/daten/private-haushalte-konsum/umweltbewusstsein-umweltverhalten#das-umweltbewusstsein-in-deutschland (Retrieved 12.02.2021)

Cajias, M., Fuerst, F., & Bienert, S. (2019). Tearing down the information barrier: the price impacts of energy efficiency ratings for buildings in the german rental market. *Energy Research & Social Science*, *47*, 177–191.

Cajias, M., & Piazolo, D. (2013). Green performs better: energy efficiency and financial return on buildings. *Journal of Corporate Real Estate*, *15*(1), 53–72.

Cespedes-Lopez, M.-F., Mora-Garcia, R.-T., Perez-Sanchez, V. R., & Perez-Sanchez, J.-C. (2019). Meta-analysis of price premiums in housing with energy performance certificates (epc). *Sustainability*, *11*(22), 1–59.

Colwell, P. F., & Dilmore, G. (1999). Who was first? an examination of an early hedonic study. *Land Economics*, *75*(4), 620–626.

Communication department of the European Commission. (2020). *Regulations, directives and other acts.* Retrieved from https://europa.eu/european-union/law/legal-acts_en (Retrieved 24.02.2021)

de Ayala, A., Galarraga, I., & Spadaro, J. V. (2016). The price of energy efficiency in the spanish housing market. *Energy Policy*, *94*, 16–24.

Eschment, A. (2020). *Wie ecosia mit werbung gutes tut.* Retrieved from https://www.zdf.de/nachrichten/digitales/ecosia-suchmaschine-sozialunternehmen-100.html (Retrieved 10.02.2021)

European Commission. (2020a). *Renovation wave: doubling the renovation rate to cut emissions, boost recovery and reduce energy poverty.* Retrieved from https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_20_1835/IP_20_1835_EN.pdf (Retrieved 10.02.2021)

European Commission. (2020b). *State of the union: Commission raises climate ambition and proposes 55% cut in emissions by 2030.* Retrieved from https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_20_1599/IP_20_1599_EN.pdf (Retrieved 10.02.2021)

Fahrmeir, L., Kneib, T., & Lang, S. (2009). *Regression modelle, methoden und anwendungen, series: Statistik und ihre anwendungen*. Springer Berlin, Heidelberg.

Federal Ministry of Finance. (2020). *Sustainable finance – what does the financial system have to do with the sustainable development goals?* Retrieved from https://www.bundesfinanzministerium.de/Content/EN/Standardartikel/Topics/Financial_markets/Articles/sustainable-finance.html# (Retrieved 24.02.2020)

Fizaine, F., Voye, P., & Baumont, C. (2018). Does the literature support a high willingness to pay for green label buildings? an answer with treatment of publication bias. *Revue d'economie politique*, *128*(5), 1013–1046.

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, *87*(417), 178–183.

Fuerst, F., Oikarinen, E., & Harjunen, O. (2016). Green signalling effects in the market for energy-efficient residential buildings. *Applied Energy*, *180*, 560–571.

Green-Towers Sustainable High-Rises GmbH. (2021). *Telephone call log (see appendix 16) between author timo deller and ceo roland weber regarding the plausibility check of data, 18.01.2021.*

Haas, G. C. (1922). *A statistical analysis of farm sales in blue earth county, minnesota, as a basis for farm land appraisal* (Tech. Rep.). Master Thesis, University of Minnesota.

Henger, R., & Voigtländer, M. (2014). Transaktions-und angebotsdaten von wohnimmobilien. eine analyse für hamburg. *IW-Trends-Vierteljahresschrift zur empirischen Wirtschaftsforschung*, *41*(4), 85–100.

Hyland, M., Lyons, R. C., & Lyons, S. (2013). The value of domestic building energy efficiency—evidence from ireland. *Energy economics*, *40*, 943–952.

International Energy Agency and the United Nations Environment Programme. (2018). *2018 global status report: towards a zero-emission, efficient and resilient buildings and construction sector.* Retrieved from https://www.worldgbc.org/sites/default/files/2018GlobalABCGlobalStatusReport.pdf (Retrieved 10.02.2021)

Janssen, C., Söderberg, B., & Zhou, J. (2001). Robust estimation of hedonic models of price and income for investment property. *Journal of Property Investment & Finance*, *19*(4), 342–360.

Kelion, L. (2020). *Apple's 2030 carbon-neutral pledge covers itself and suppliers.* Retrieved from https://www.bbc.com/news/technology-53485560 (Retrieved 10.02.2021)

Kholodilin, K. A., Mense, A., & Michelsen, C. (2017). The market value of energy efficiency in buildings and the mode of tenure. *Urban Studies*, *54*(14), 3218–3238.

Kim, S., Lim, B. T. H., & Kim, J. (2016). *Green features, symbolic values and rental premium: systematic review and meta-analysis, conference paper, unsw australia.*

Knight, J. R. (2002). Listing price, time on market, and ultimate selling price: Causes and effects of listing price changes. *Real estate economics*, *30*(2), 213–237.

KPMG Luxembourg. (2019). *European responsible investing fund market 2019.* Retrieved from https://assets.kpmg/content/dam/kpmg/lu/pdf/lu-en-EuropeanResponsibleInvestmentFund-2019.pdf (Retrieved 15.02.2021)

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy*, *74*(2), 132–157.

Leavens, D. H. (1939). Report of the detroit meeting, december 27-30, 1938. *Econometrica*, *7*(2), 167–190.

Leopoldsberger, G., Bienert, S., Brunauer, W., Bobsin, K., & Schützenhofer, C. (2011). Energising property valuation: Putting a value on energy-efficient buildings. *Appraisal Journal*, *79*(2), 115–125.

Ma, J., Sheren, M., & Zadek, S. (2016). *G20 green finance synthesis report.* Retrieved from http://unepinquiry.org/wp-content/uploads/2016/09/Synthesis_Report_Full_EN.pdf (Retrieved 10.02.2021)

Manus, C. (2020). *Wohnungsmarkt bleibt angespannt.* Retrieved from https://www.fr.de/frankfurt/wohnungsmarkt-in-frankfurt-und-rhein-main-bleibt-angespannt-90089517.html (Retrieved 01.03.2021)

Markowitz, H. M. (1991). Foundations of portfolio theory. *The journal of finance*, *46*(2), 469–477.

Morgan, C. (2020). *How adidas is turning plastic ocean waste into sneakers and sportswear.* Retrieved from https://www.businessinsider.com/adidas-sneakers-plastic-bottles-ocean-waste-recycle-pollution-2019-8 (Retrieved 02.01.2021)

Real Estate Pilot AG. (2020). *Alle objekte eine seite besuchen, alle objekte finden.* Retrieved from https://geomap.immo/alle-objekte/ (Retrieved 02.03.2021)

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, *82*(1), 34–55.

Schleidt, D. (2020). *Frankfurt sieht sich als brexit-gewinner neue arbeitsplätze.* Retrieved from https://www.faz.net/aktuell/rhein-main/neue-arbeitsplaetze-frankfurt-sieht

-sich-als-brexit-gewinner-16828637.html (Retrieved 15.02.2021)

Sopranzetti, B. J. (2010). Hedonic regression analysis in real estate markets: a primer. In *Handbook of quantitative finance and risk management* (pp. 1201–1207). Springer.

Statista Research Department. (2020). *Günstigste städte für eigentumswohnungen in deutschland im 4. quartal 2017.* Retrieved from https://de.statista.com/statistik/daten/studie/201808/umfrage/die-guenstigsten-staedte-fuer-eigentumswohnungen-in-deutschland/#professional (Retrieved 23.02.2021)

Statista Research Department. (2021). *Einwohnerzahl der kreisfreien städte und landkreise der metropolregion frankfurtrheinmain im jahr 2019.* Retrieved from https://de.statista.com/statistik/daten/studie/1185422/umfrage/bevoelkerung-in-metropolregion-frankfurt-rhein-main-kreise/ (Retrieved 07.01.2021)

Statistisches Bundesamt. (2020). *Wohnen – eigentümerquote nach bundesländern.* Retrieved from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Wohnen/Tabellen/eigentuemerquote-nach-bundeslaender.html (Retrieved 13.02.2021)

Urban, D., & Mayerl, J. (2018). *Angewandte regressionsanalyse: Theorie, technik und praxis*. Springer, Wiesbaden.

Verbraucherzentrale Nordrhein-Westfalen, & Verbraucherzentrale Rheinland-Pfalz. (2020). *Energieausweis: Was sagt dieser steckbrief für wohngebäude aus?* Retrieved from https://www.verbraucherzentrale.de/wissen/energie/energetische-sanierung/energieausweis-was-sagt-dieser-steckbrief-fuer-wohngebaeude-aus-24074 (Retrieved 12.02.2021)

Wahlström, M. H. (2016). Doing good but not that well? a dilemma for energy conserving homeowners. *Energy Economics*, *60*, 197–205.

Wallace, H. A. (1926). Comparative farm-land values in iowa. *The Journal of Land & Public Utility Economics*, *2*(4), 385–392.

Williams, J. B. (1938). The theory of investment value, amsterdam.

Junior Management Science 7(3) (2022) 826-873

Junior Management Science

journal homepage: www.jums.academy

Corporate Social Responsibility and the Effects of Sustainable Corporate Practices and Various Greenwashing Methods on Corporate Reputation

Corporate Social Responsibility und die Effekte nachhaltiger Unternehmenspraktiken und unterschiedlicher Greenwashing-Methoden auf die Unternehmensreputation

Simon Mütze

Leibniz Universität Hannover

Abstract

As awareness of sustainable consumption continues to grow, corporate social responsibility (CSR) is becoming increasingly relevant. However, when it comes to communicating corporate sustainability (green marketing), false or vague claims, so-called greenwashing, have now reached epidemic dimensions. For this reason, many consumers have built up a fundamental skepticism towards this type of communication. The purpose of this study is to help marketers establish and communicate effective as well as sincere corporate practices to address the challenges posed by greenwashing. Therefore, this study uses a field experiment to investigate how often greenwashing is recognized by consumers in the first place, what companies need to consider for successful CSR and, in particular, what effects different CSR initiatives, greenwashing methods and greenwashing scandals have on corporate reputation. One of the key findings is that greenwashing is strongly condemned and has a strong negative impact on corporate reputation, but due to high consumer skepticism and confusion, as well as low environmental knowledge, it is hardly distinguished from actual CSR measures. Proposed solutions include awareness campaigns for a better understanding of environmental claims in advertising, and government reform and monitoring of eco-labels.

Zusammenfassung

Mit dem zunehmenden Bewusstsein für nachhaltigen Konsum gewinnt Corporate Social Responsibility (CSR) immer mehr an Bedeutung. Bei der Kommunikation ebendieser unternehmerischen Nachhaltigkeit (Green Marketing) haben falsche oder vage Behauptungen, das sogenannte Greenwashing, jedoch mittlerweile epidemische Ausmaße angenommen. Aus diesem Grund haben viele Verbraucher eine grundsätzliche Skepsis gegenüber dieser Art von Kommunikation aufgebaut. Ziel dieser Studie ist es, Marketern dabei zu helfen, effektive und aufrichtige Unternehmenspraktiken zu etablieren und zu kommunizieren, um den Herausforderungen des Greenwashings zu begegnen. In dieser Studie wird daher mittels eines Feldexperiments untersucht, wie häufig Greenwashing überhaupt von Verbrauchern erkannt wird, was Unternehmen für eine erfolgreiche CSR beachten müssen und insbesondere, welche Auswirkungen unterschiedliche CSR-Initiativen, Greenwashing-Methoden und Greenwashing-Skandale auf die Unternehmensreputation haben. Zu den zentralen Erkenntnissen zählt, dass Greenwashing zwar stark verurteilt wird und sich stark negativ auf die Unternehmensreputation auswirkt, aufgrund der hohen Verbraucherskepsis und -verwirrung sowie des geringen Umweltwissens jedoch kaum von tatsächlichen CSR-Maßnahmen unterschieden wird. Lösungsvorschläge umfassen Sensibilisierungskampagnen für ein besseres Verständnis von Umweltaussagen in der Werbung sowie eine staatliche Reform und Überwachung von Umweltsiegeln.

Keywords: CSR; Greenwashing; Unternehmensreputation; Unternehmerische Nachhaltigkeit; Umweltsiegel.

DOI: https://doi.org/10.5282/jums/v7i3pp826-873

1. Einleitung

Corporate Social Responsibility (CSR) beschreibt die soziale Verantwortung von Unternehmen und ist eines der Schlüsselelemente in der heutigen strategischen Unternehmensführung (vgl. Baumgartner, 2014, S. 258; Lewis, 2003, S. 363-364). Der Ausgangspunkt eines erfolgreichen Unternehmens ist nicht das Unternehmen selbst, sondern der Markt. Um im Wettbewerb bestehen zu können, muss ein Unternehmen seine Vision und Mission an die Bedürfnisse und Wünsche des Marktes anpassen (vgl. Kok & Biemans, 2009, S. 517). Dabei hat die wachsende Besorgnis über die globale Erwärmung das Bewusstsein der Konsumenten für Umweltbelange stark erhöht (vgl. Y.-S. Chen, 2008, S. 531-533). Sie verlangen nach Produkten, die nachhaltig und vorteilhaft für die Umwelt sind. Noch im Jahr 2009 ist der Verbrauchermarkt für umweltfreundliche Produkte und Dienstleistungen auf $230 Milliarden geschätzt worden, 2015 bereits auf $845 Milliarden (vgl. Delmas & Burbano, 2011, S. 64).

Aufgrund dieses zunehmenden Interesses an umweltfreundlichen Produkten und Dienstleistungen haben Unternehmen beträchtliche Investitionen in die Herstellung und Vermarktung sogenannter „grüner“ Produkte getätigt, um von diesen expandierenden Märkten zu profitieren (vgl. Bhatia & Jain, 2013, S. 1-3). Die zunehmenden Umweltbedenken und der damit einhergehende Druck, sich umweltbewusst zu verhalten, haben das Umweltmanagement an die Spitze der Agenden vieler Unternehmen gerückt (vgl. A. King & Lenox, 2002, S. 297). Folglich werden stetig mehr umweltfreundliche Praktiken angewendet, die nicht nur das Endprodukt selbst betreffen, sondern auch die gesamte Produktion und Bereitstellung (vgl. Kivimaa & Kautto, 2010, S. 290). So ist bspw. allein zwischen 2009 und 2010 das Angebot an umweltfreundlichen Produkten um 73% gestiegen (vgl. TerraChoice, 2010, S. 6). Um Konsumenten mit einem umweltbewussten Lebensstil anzusprechen, kreieren Unternehmen Werbung mit umweltfreundlichen Behauptungen (vgl. Divine & Lepisto, 2005, S. 275). Diese Kommunikation wird Green Marketing genannt und gilt als das effizienteste Konzept, um auf die Bedürfnisse und Wünsche des Marktes zu reagieren (vgl. Aji & Sutikno, 2015, S. 434). Aus diesem Grund hat auch die Anzahl der „grünen“ Werbekampagnen im 21. Jahrhundert exponentiell zugenommen. Bereits im Jahr 2009 verfügten mehr als 75% der S&P-500-Unternehmen über Website-Bereiche, die der Offenlegung ihrer Umwelt- und Sozialpolitik gewidmet waren (vgl. Alves, 2009, S. 8; Delmas & Burbano, 2011, S. 64).

Allerdings spiegeln nicht alle Green Marketing-Behauptungen das Umweltverhalten der Unternehmen korrekt wider. Während sich einige Unternehmen tatsächlich stark für mehr Nachhaltigkeit engagieren, überzeichnen andere ihre Bemühungen, oder behaupten, umweltbewusst zu sein, obwohl dies nicht der Fall ist (vgl. L. Mitchell & Ramey, 2011, S. 41). Dieses Phänomen wird Greenwashing genannt. Beim Greenwashing verbreiten Unternehmen oder andere Organisationen gezielt Desinformationen, um in der Öffentlichkeit ein umweltverantwortliches Bild zu vermitteln (vgl. Concise Oxford English Dictionary, 2010, o. S.). Zu den bekannten Beispielen von Unternehmen, die des Greenwashings beschuldigt werden, gehört bspw. das Mineralölunternehmen British Petroleum (BP), das eine auffällige grüne Kampagne mit suggestivem grünem Logo eingeführt hat, während es weniger als 1% des Umsatzes mit erneuerbaren Energien erzielt (vgl. BUND, 2016). Weiterhin geriet auch Coca-Cola in die Kritik, als das Unternehmen die nicht nachprüfbare Behauptung aufstellte, es habe seinen Wasserverbrauch um etwa 4% pro Jahr gesenkt, um die Umwelt weniger stark zu belasten (vgl. Lyon & Montgomery, 2015, S. 224). Als weiteres Beispiel hat SC Johnson eine Sammelklage gegen eine ungenannte Summe beigelegt, nachdem sich Konsumenten getäuscht fühlten, weil das „Greenlist“-Label auf mehreren Produkten von SC Johnson darauf schließen ließ, die Produkte seien von einer unabhängigen dritten Partei zertifiziert. Tatsächlich ist „Greenlist“ ein Label, welches von SC Johnson selbst entworfen und verwaltet wurde (vgl. Hoffman, 2013, o. S.).

Beispiele wie diese sind dabei keine Seltenheit; ganz im Gegenteil: Bei über 95 Prozent der von TerraChoice (2010, S. 16) betrachteten als umweltfreundlich deklarierten Produkte konnte eine Form von Greenwashing nachgewiesen werden. Auch die renommierte Werbeagentur Ogilvy & Mather erklärt, dass Greenwashing „epidemische Ausmaße“ (Hsu, 2011) erreicht habe. Diese sprunghaft ansteigende Verbreitung von Greenwashing kann schwerwiegende negative Auswirkungen auf das Vertrauen in umweltfreundliche Produkte haben. Wenn Konsumenten den Behauptungen von Unternehmen nicht mehr glauben können und nicht dazu in der Lage sind, tatsächliches CSR eindeutig von Greenwashing zu unterscheiden, kann dies den gesamten Markt für umweltfreundliche Produkte und Dienstleistungen gefährden (vgl. Delmas & Burbano, 2011, S. 64). Aus diesem Grund nimmt die akademische Aufmerksamkeit für das Phänomen Greenwashing rasant zu (vgl. Lyon & Montgomery, 2015, S. 223-224). In Anbetracht der Prävalenz und des potenziellen Ausmaßes von Greenwashing ist die empirische Forschung zu seinen Wirkungen allerdings begrenzt. Folglich sind viele Fragen unbeantwortet, wie bspw. zu den Effekten unterschiedlicher Greenwashing-Methoden auf die Einstellung der Konsumenten gegenüber einer Marke. Von den verschiedenen Aspekten der Kundeneinstellung und -perspektive ist die wahrgenommene Unternehmensreputation eine der wichtigsten Bedingungen für den Erfolg eines Unternehmens auf wettbewerbsorientierten Märkten (vgl. Keh & Xie, 2009, S. 732; Park 2019, S. 215). Die Unternehmensreputation geht dabei über das Unternehmensimage hinaus. Sie umfasst nicht nur das aktuelle Bild, das die Stakeholder von einem Unternehmen haben, sondern auch die in der Vergangenheit aufgebauten und für die Zukunft relevanten Unterstützungspotentiale wie bspw. Vertrauen oder Interesse (vgl. Wiedmann, Fombrun & van Riel, 2006, S. 99).

In dieser Studie soll daher untersucht werden, wie häufig Greenwashing überhaupt von den Konsumenten erkannt wird, was Unternehmen für erfolgreiches CSR beachten müssen und insbesondere welche Effekte unterschiedliche CSR-

Initiativen, Greenwashing-Methoden und Greenwashing-Skandale auf die Unternehmensreputation haben.

Im folgenden zweiten Kapitel werden zunächst die wichtigsten theoretischen Grundlagen erläutert. Darauf aufbauend wird in Kapitel drei der konzeptionelle Bezugsrahmen dieser Studie entwickelt und mithilfe von Ergebnissen aus der Literatur Hypothesen abgeleitet. Anschließend wird im vierten Kapitel das durchgeführte Experiment beschrieben, analysiert und ausgewertet. Anhand der erzielten Ergebnisse werden in Kapitel fünf Implikationen für die Marketingpraxis und -forschung abgeleitet sowie die Limitationen dargestellt. Abschließend folgen im sechsten Kapitel das Fazit und ein Ausblick.

2. Grundlagen

2.1. Corporate Social Responsibility (CSR)

Das Hauptziel eines jeden Unternehmens ist die Schaffung von Werten für seine Stakeholder (vgl. Dembek, York & Singh, 2018, S. 1600; Jermsittiparsert, Siam, Issa, Ahmed & Pahi, 2019, S. 741). Auch wenn gewinnorientierte Unternehmen keine gemeinnützigen Organisationen sind, ist ihr Interesse gelegentlich auf die Unterstützung gemeinnütziger Aktivitäten gerichtet. Laut McWilliams und Siegel (2001, S. 117) wird CSR als solche Maßnahmen definiert, in denen Unternehmen nicht nur die geltenden Vorschriften einhalten, sondern Aktionen durchführen, die offenbar ein soziales Gut fördern, welches über die Interessen des Unternehmens und das gesetzlich vorgeschriebene Maß hinausgeht. CSR ist demnach ein Teil der wirtschaftlichen, rechtlichen und moralischen Aktivitäten eines Unternehmens, das darauf ausgerichtet ist, nicht nur ökonomische, sondern auch gesellschaftlich relevante Aspekte zu unterstützen (vgl. Salvioni & Gennari, 2014, S. 469; Enobong, 2017, S. 82).

Dies ist jedoch nur eine Interpretation. Mittlerweile existieren zahlreiche Definitionen von CSR, was die theoretische Entwicklung und Messung schwierig macht. CSR-Aktivitäten sind entwickelt worden, um gesellschaftlich vorteilhafte Charakteristika in Produkte und Herstellungsprozesse einzubeziehen (z.B. Papier aus nachhaltiger Forstwirtschaft), fortschrittliche Praktiken des Personalmanagements zu übernehmen (z.B. Förderung der Mitbestimmung von Mitarbeitern) oder durch Recycling und Verschmutzungsreduzierung ein höheres Niveau der Umweltleistung zu erreichen. Auch wenn der Fokus dieser Aktivitäten auf gesellschaftlichen Verbesserungen liegt, sollen letztlich sowohl die Allgemeinheit als auch das Unternehmen selbst von diesen Initiativen profitieren (vgl. McWilliams, Siegel & Wright, 2006, S. 1-2).

Es gibt diverse empirisch belegte strategische Motive dafür, warum Unternehmen umweltfreundliche oder soziale Aktivitäten betreiben sollten. Dazu gehören: Einen Wettbewerbsvorteil in der eigenen Branche erlangen, sich einen Ruf der sozialen Verantwortung erarbeiten oder auch die Reduzierung der Betriebskosten (vgl. Bansal & Roth, 2000, S. 732; Lyon & Maxwell, 2008, S. 252-256; Porter & van der Linde, 1995, S. 130-134). Moderne Organisationen müssen nicht nur über den Preis und die Qualität ihrer Produkte und Dienstleistungen konkurrieren, sondern auch über die Umweltfreundlichkeit. Die öffentliche Wahrnehmung der Bemühungen von Unternehmen, ehrlich und sozial verantwortlich zu handeln, steht in direktem Zusammenhang mit der Erzielung und Aufrechterhaltung von Wettbewerbsvorteilen. Da Unternehmen diesen Nutzen eines ökologischen bzw. sozialen Images erkannt haben, erstellen viele von ihnen CSR-Programme, um ihre Gewissenhaftigkeit gegenüber der Gesellschaft zu demonstrieren. (vgl. Lyon & Maxwell, 2008, S. 252-256). Aufgrund asymmetrischer Informationen ist es allerdings schwierig, die genauen Motivationen und Wirkungen von CSR zu untersuchen. Manager erkennen bspw., dass viele externe Stakeholder die CSR-Aktivität positiver beurteilen, wenn sie ohne ökonomische Hintergedanken getroffen wurde. Dies verleitet sie dazu, die eher strategischen Motivationen (wie Produktförderung, Kostenkontrolle und Reputationsbildung) hinter ihren CSR-Aktivitäten nicht offen zu legen, insbesondere nicht in Unternehmens-publikationen wie Geschäftsberichten. Diese Situation der asymmetrischen Informationen erschwert es häufig, die verschiedenen strategischen und soziale Motivationen für CSR unterscheiden zu können (vgl. McWilliams et al., 2006, S. 9).

Baron (2001, S. 11) bezeichnet die Verwendung von CSR zur Gewinnerzielung als strategisches CSR und weist darauf hin, dass es die Motivation für das Handeln ist, die soziales, verantwortungsbewusstes, von egoistischem Handeln trennt. Das heißt, wenn die Motivation darin besteht, der Gesellschaft auf Kosten von Profiten zu dienen, ist die Handlung sozial verantwortlich, wenn aber die Motivation darin besteht, das Geschäftsergebnis zu verbessern, dann ist die Handlung strategisch.

Ein wichtiger Beitrag zur Debatte über das Verhalten von Unternehmen kam von dem Konzept der „Triple Bottom Line“, das von John Elkington als ein Rahmenwerk für Nachhaltigkeit konzipiert worden ist und die sozialen, ökologischen und ökonomischen Auswirkungen eines Unternehmens ausbalanciert. Elkington (1998, S. 37-51) erklärt, dass der Weg zu einer herausragenden Triple-Bottom-Line-Leistung über effektive und langfristige Partnerschaften zwischen Unternehmen und den Stakeholdern führt. Auch wenn das Triple-Bottom-Line-Konzept bereits in den späten 1990er Jahren als praktischer Ansatz für Nachhaltigkeit populär wurde ist es in der CSR-Diskussion nach wie vor relevant, weil es darauf hinweist, dass Unternehmen ein sozial und ökologisch verantwortungsbewusstes Verhalten an den Tag legen müssen, welches positiv mit ihren wirtschaftlichen Zielen in Einklang gebracht werden kann (vgl. Adidas, 2019, S. 9; Van Marrewijk, 2003, S. 8-9). Die drei Aspekte der Nachhaltigkeit (ökonomisch, ökologisch und sozial) können in einen Ansatz übersetzt werden, mit dem sich Unternehmen befassen müssen. Die Abbildung 1 zeigt das Verhältnis von sozialer, ökologischer und ökonomischer Verantwortung. Das Modell setzt sich aus drei Säulen zusammen, die gemeinsam die Grundlage für CSR, bzw. unternehmerische Nachhaltigkeit bilden:

- People: Die soziale Dimension. Sie umfasst eine Vielzahl von Aspekten, die die Auswirkungen der Unternehmenstätigkeit auf die Menschen innerhalb und außerhalb der Organisation betreffen, wie bspw. gute Arbeitsbedingungen sowie Gesundheit und Sicherheit.
- Planet: Die ökologische Dimension. Diese Dimension bezieht sich auf die Auswirkungen der Unternehmenstätigkeit auf die natürliche Umwelt.
- Profit: Die ökonomische Dimension. Diese Dimension bezieht sich auf die Wertschöpfung durch die Produktion von Waren und Dienstleistungen sowie durch die Schaffung von Arbeitsplätzen und Verdienstmöglichkeiten (vgl. Graafland & Van de Ven, 2006, S. 113).

Abbildung 1: Das Triple-Bottom-Line Modell

Quelle: Eigene Darstellung in Anlehnung an van Marrewijk 2003, S. 101

CSR kann auch bei strategischer Motivation gesellschaftliche Vorteile bringen. Das Problem ist, dass Unternehmen hier oftmals die ökonomischen Aspekte in das Zentrum ihrer Bemühungen stellen und die sozialen Aspekte lediglich Mittel zum Zweck sind, was nicht dem Grundgedanken von CSR entspricht. Es ist wichtig, dass Unternehmen nicht nur die ökonomischen, sondern auch die sozialen und ökologischen Vorteile als wertschöpfend erkennen und sich aktiv um diese bemühen. Richardson, Welker und Hutchinson (1999, S. 37-38) sehen CSR als eine fortschreitende Betonung und Verpflichtung der Unternehmen auf die moralischen Komponenten ihrer Verwaltungs- und Organisationsstruktur. Heute haben viele Organisationen CSR in ihre Unternehmensstrategie integriert, um ihre Kunden zufriedenzustellen (vgl. S. Lee, Lee, Gao, Xiao & Conklin, 2018, S. 356; Mohammed & Rashid, 2018, S. 361).

2.1.1. Geschichte und Entwicklung von CSR

Die Entwicklung von CSR ist gewiss ein stetiger Prozess, der nicht auf einzelne Ereignisse reduziert werden kann. In diesem Abschnitt sollen dennoch die wichtigsten Meilensteine erläutert werden, um das Ausmaß und die Bedeutung von CSR sowohl für Unternehmen als auch für die Gesellschaft besser verstehen zu können.

Die Geschichte beginnt Anfang der 1950er Jahre, als H. R. Bowen (1953, S. 17) erklärte, dass die großen Konzerne jener Zeit bedeutende Macht konzentrieren und ihr Handeln spürbare Auswirkungen auf die Gesellschaft habe. Als Ergebnis seiner Überzeugung legte er die Idee dar, spezifische Prinzipien zu definieren, mit denen Unternehmen ihre soziale Verantwortung erfüllen können. Er erkannte, dass die Entscheidungen und Handlungen von Managern eine ganze Reihe von Stakeholdern beeinflussen, was wiederum einen direkten Einfluss auf die Qualität der Gesellschaft als Ganzes hat (vgl. H. R. Bowen, 1953, S. 3). Wie Carroll (2008, S. 5-6) erklärt, war Bowen mit seinem neuen Managementansatz seiner Zeit weit voraus. Sein Ansatz war die erste akademische Arbeit, die sich speziell auf die Lehre der sozialen Verantwortung konzentrierte, was ihn zum „Vater der CSR“ (Carroll 1999, S. 270) macht.

In den 1960er Jahren entwickelte sich die Gesellschaft weiter und eine wachsende Protestkultur entstand, die sich hauptsächlich um Bürgerrechte und Antikriegsproteste drehte (vgl. Waterhouse, 2017, S. 15). Dementsprechend näherten sich Wissenschaftler der CSR als eine Antwort auf die Probleme und Wünsche der neuen, modernen Gesellschaft. Ein bemerkenswertes Beispiel aus dieser Zeit ist Keith K. Davis (1960, S. 73), der erklärte, dass die sich gegenwärtig vollziehenden wichtigen sozialen, wirtschaftlichen und politischen Veränderungen für Geschäftsleute einen gewissen Zwang darstellen, ihre Rolle in der Gesellschaft und ihre soziale Verantwortung zu überdenken.

Trotz der zunehmenden gesellschaftlichen Entwicklungen hin zu mehr sozialer Verantwortung gab es auch Kritiker des CSR-Ansatzes. Insbesondere Milton Friedman, späterer Nobelpreisträger für Wirtschaftswissenschaften (1976), unterstrich 1962 die Rolle der Unternehmen im freien kapitalistischen System, in dem sich die Unternehmen auf das Streben nach ökologischen Vorteilen beschränken sollten (vgl. Friedman, 1962). Friedman hat diesen Ansatz in dem Artikel „The Social Responsibility of Business is to Increase its Profits“ (Friedman, 1970, S. 32-33) weiter untersucht, in dem er CSR-Aktivitäten als eine unangemessene Nutzung der Unternehmensressourcen betrachtet, die zu einer ungerechtfertigten Ausgabe von Geld für das allgemeine soziale Wohl führen würde.

In den 1970er Jahren entstanden einige der heute, im Hinblick auf die soziale Verantwortung, bekanntesten Unternehmen. Zu nennen sind hier beispielsweise der 1976 im Vereinigten Königreich gegründete The Body Shop sowie das 1978 in den USA gegründete Unternehmen Ben & Jerry's. Ob als Antwort auf die neuen sozialen Erwartungen oder aufgrund einer First-Mover-Strategie, dies sind zwei bemerkenswerte Beispiele für Unternehmen, die begannen, sich mit den sozialen und öffentlichen Fragen der Zeit zu befassen (vgl. Latapí Agudelo, Jóhannsdóttir & Davídsdóttir, 2019, S. 6). Der Erfolg dieser Unternehmen hat dazu beigetragen, dass der Begriff CSR immer populärer wurde, was weiterhin dazu führte, dass er in vielen verschiedenen Bereichen und in einem solchen Ausmaß verwendet worden ist, dass seine Bedeutung unklar wurde und infolgedessen für jeden etwas an-

deres bedeutete (vgl. Sethi, 1975, S. 58; Votaw, 1972, S. 25).

Diese Unsicherheit in der Bedeutung dauerte bis 1979 an, als Carroll die wohl erste einheitliche Definition des Begriffs der CSR vorschlug. Laut ihm umfasst die soziale Verantwortung von Unternehmen die wirtschaftlichen, rechtlichen, ethischen und diskretionären Erwartungen, die die Gesellschaft zu einem bestimmten Zeitpunkt an Organisationen stellt (vgl. Carroll, 1979, S. 500).

In den 1980er Jahren führten einige gesellschaftliche Ereignisse zur Weiterentwicklung der CSR. Zu den wichtigsten gehören: Die Gründung der Generaldirektion Umwelt der Europäischen Kommission (1981), die Einrichtung der Weltkommission für Umwelt und Entwicklung (1983), die Atomkatastrophe von Tschernobyl (1986), und die Gründung des Weltklimarats (1988). Auch wenn sich diese Ereignisse nicht unmittelbar auf CSR bezogen und daher keinen direkten Einfluss auf die Entwicklung des Konzepts hatten, spiegelten sie doch ein wachsendes Bewusstsein der internationalen Gemeinschaft in Bezug auf Umweltschutz und nachhaltige Entwicklung und indirekt auf das Verhalten von Unternehmen wider (vgl. Latapí Agudelo et al., 2019, S. 7).

Als einen der wichtigsten Beiträge zur CSR präsentierte Carroll 1991 die „Pyramide der sozialen Verantwortung der Unternehmen“ (Abb. 2) mit dem Ziel, den Führungskräften, die ihre Verpflichtungen gegenüber den Aktionären mit ihren Verpflichtungen gegenüber einem breiteren Kreis von Stakeholdern in Einklang bringen mussten, einen nützlichen CSR-Ansatz bereitzustellen. Mit der Pyramide der CSR repräsentierte Carroll das, was er als die vier Hauptaufgaben eines jeden Unternehmens definierte: 1. Die wirtschaftliche Verantwortung, die die Grundlage für die anderen Ebenen der Pyramide bildet; 2. Die rechtliche Verantwortung des Unternehmens; 3. Die ethische Verantwortung, die das Verhalten des Unternehmens über die gesetzestreuen Pflichten hinaus prägt; und 4. Die philanthropische Verantwortung des Unternehmens im Hinblick auf dessen Beitrag zur Verbesserung der Lebensqualität der Gesellschaft (vgl. Carroll et al., 1991, S. 42).

Abbildung 2: Die Pyramide der sozialen Verantwortung von Unternehmen

Quelle: Eigene Darstellung in Anlehnung an Carroll et al. (1991), S. 42

Der Globalisierungsprozess der 1990er Jahre steigerte die globale Reichweite multinationaler Unternehmen und der Kapitalismus weitete sich rasch aus. Dies führte dazu, dass sich die Bedenken vieler Unternehmen hinsichtlich ihrer Wettbewerbsfähigkeit, Reputation und globaler Sichtbarkeit vergrößerten (vgl. Carroll, 2015, S. 88). So kamen Burke und Logsdon (1996) auf die Idee, nach Beweisen zu suchen, dass CSR zu einer Verbesserung der finanziellen Leistung von Unternehmen führt, womit sie wohl die ersten waren, die den strategischen Nutzen von CSR untersuchten. Sie erkannten, dass CSR strategisch eingesetzt werden kann, um dadurch die Effektivität eines Unternehmens bei der Erreichung seiner Hauptziele zu verbessern und die Wettbewerbsfähigkeit zu erhöhen (vgl. Burke & Logsdon, 1996, S. 495).

Die weltweite Bekanntheit von CSR wurde auch durch internationale Zertifizierungen beeinflusst, die sich mit der sozialen Verantwortung befassen. Die ISO 26000 dient als Leitfaden für die Art und Weise, wie Unternehmen sozial verantwortlich handeln können. Für die CSR-Bewegung ist sie von besonderer Bedeutung, weil sie von 450 Experten aus 99 Ländern entwickelt wurde und bisher von mehr als 80 Ländern als Richtlinie für nationale Standards übernommen wurde (vgl. Schwartz & Tilling, 2009, S. 290).

Das Verständnis von CSR hat sich von einer persönlichen Entscheidung von Geschäftsleuten in den 1950er Jahren zu einem Entscheidungsprozess in den 1980er Jahren und zu einer strategischen Notwendigkeit zu Beginn der 2000er Jahre entwickelt. Insbesondere hat sich auch der Zweck der Existenz von Unternehmen von der Beschränkung auf die Erzielung wirtschaftlicher Gewinne in den 1950er und 60er Jahren hin zu der in den 2010er Jahren geäußerten Überzeugung entwickelt, dass die Aufgabe von Unternehmen darin bestehen sollte, gemeinsame Werte zu schaffen. Auf diesem Weg wurde CSR durch zahllose wissenschaftliche Veröffentlichungen sowie Regierungsentscheidungen und soziale sowie internationale Bewegungen beeinflusst. Es lässt sich festhalten, dass das Verständnis von Corporate Social Responsibility dynamisch ist und auf soziale Erwartungen an das Verhalten von Unternehmen reagiert (vgl. Latapí Agudelo et al., 2019, S. 16).

Heute ist CSR unverzichtbar geworden und ein wesentlicher Baustein in der strategischen Unternehmensführung (vgl. Baumgartner, 2014, S. 269). Die Frage, wie mithilfe von CSR bestmöglich sowohl wirtschaftliche als auch soziale und ökologische Ziele erreicht werden können, ist zu einem der am meisten beachteten Themen in diesem Forschungsbereich geworden (vgl. Costa & Menichini, 2013, S. 150).

2.2. Green Marketing

Da sich Konsumenten zunehmend der potenziellen Umweltauswirkungen ihrer Einkäufe bewusstwerden, sind Produkte, die als „umweltfreundlich“, „organisch“ oder „nachhaltig“ gekennzeichnet sind beliebter denn je (vgl. Baum, 2012, S. 423; Parguel, Benoit-Moreau & Russell, 2015, S. 107). Unternehmen suchen ständig nach neuen Wegen, ihre Produkte im Zeitalter des Umweltschutzes zu differenzieren. Als Reaktion auf die wachsende Nachfrage der Konsumenten nach umweltfreundlichen Produkten sind Werbeaussagen, die genau das versprechen, zu einem wichtigen Bestandteil der Vermarktung geworden (vgl. Y.-S. Chen & Chang,

2013, S. 489; Segev, Fernandes & Hong, 2016, S. 91). Dieser Ansatz wird Green Marketing genannt. Aufgrund der zunehmenden Bedeutsamkeit ist die Auseinandersetzung mit diesem Trend mittlerweile zudem ein vielbeachtetes Thema in der akademischen Forschung (vgl. Taylor, 2015, S. 573).

Wissenschaftler haben viele andere Begriffe für Green Marketing definiert, wie bspw. „ökologisches Marketing", „Umweltmarketing" oder „verantwortungsbewusstes Marketing". Alle Begriffe haben jedoch einen gemeinsamen Fokus: Die Berücksichtigung und Minimierung von Umweltschäden (vgl. Polonsky, 2011, S. 1311). Es beschreibt ein Konzept bzw. eine Strategie, die ein Unternehmen anwendet, um für seine umweltfreundlichen Methoden als Zeichen seiner Bemühungen um Umweltbelange zu werben. Manju (2012, S. 40-41) beschreibt es als Marketing für Produkte und Dienstleistungen, bei denen die Produktion, der Konsum und die Entsorgung von Produkten und Dienstleistungen auf eine weniger umweltschädliche Weise erfolgen.

Da globale Probleme wie bspw. der Klimawandel oder die langfristige Entsorgung von Plastik die Wahrnehmung und den Lebensstil der Konsumenten stark beeinflussen, müssen Unternehmen sorgfältig für ihre umweltfreundlichen Produkte werben, damit die Kunden nicht den Eindruck bekommen, dass sie durch irreführende Werbung getäuscht werden (vgl. Aji & Sutikno, 2015, S. 436). Die Mehrdeutigkeit von Begriffen wie „grün" oder „ozonfreundlich" führt häufig zu für den Konsumenten sehr undurchsichtigen Situationen (vgl. Newell, Goldsmith & Banzhaf, 1998, S. 48). Grüne Behauptungen sollten klar, wahr und genau sein. Dennoch sind viele Umweltaussagen zweideutig und irreführend (vgl. Y.-S. Chen & Chang, 2013, S. 489). Ist dies der Fall, wird von Greenwashing gesprochen.

2.3. Greenwashing

Mit der zunehmenden Besorgnis über die globale Erwärmung wächst auch die Beachtung der aktuellen Umweltproblematik (vgl. Y.-S. Chen, 2008, S. 531). Konsumenten erwarten von den Unternehmen stets neue Produkte, die nicht nur gut für sie selbst sind, sondern auch für die Umwelt. Aus diesem Grund ist die Nachhaltigkeit von Produkten und Dienstleistungen und ein daraus entstehendes nachhaltiges Unternehmensbild zu einem wesentlichen Wettbewerbsvorteil gegenüber konkurrierenden Unternehmen geworden. Um diese Nachfrage der Konsumenten bedienen zu können, lassen sich immer mehr Unternehmen zu Greenwashing-Aktivitäten verleiten (vgl. Parguel, Benoît-Moreau & Larceneux, 2011, S. 15; Laufer, 2003, S. 255-257).

Als Ausdruck für die Kombination der Begriffe „green" und „brainwashing" wird Green-washing definiert als Desinformation, die von einer Organisation verbreitet wird, um in der Öffentlichkeit ein umweltverantwortliches Bild zu vermitteln (vgl. Concise Oxford English Dictionary, 2010, o. S.) bzw. als ein Akt der Irreführung der Konsumenten hinsichtlich der Umweltpraktiken eines Unternehmens oder des Umweltnutzens eines Produkts (vgl. TerraChoice, 2010, o. S.). Im Wesentlichen handelt es sich beim Greenwashing um die absichtliche Handlung eines Unternehmens, potenziell schädliche Informationen zu verschleiern oder Informationen zu verbreiten, die ein falsches Bild davon vermitteln, wie umweltfreundlich das Unternehmen ist (vgl. L. Mitchell & Ramey, 2011, S. 41; Furlow, 2010, S. 23-24).

Greenwashing ist dabei eine weit verbreitete Methode zur Irreführung von Konsumenten. Von den mehr als 5.000 selbst deklarierten grünen Produkten, die von TerraChoice (2010, S.16) überprüft worden sind, weisen 95% eine Form von Greenwashing auf. Eines der populärsten Beispiele für Firmen, die Greenwashing betreiben, ist das Mineralöl- und Erdgasunternehmen Shell. Shell gibt an, signifikant in Windenergie und weitere erneuerbare Energiequellen zu investieren, ohne genaue Zahlen zu nennen. Auf der Homepage finden sich zahlreiche Behauptungen bezüglich der Umweltfreundlichkeit des Unternehmens (vgl. Shell, 2021) und Werbeanzeigen in zahlreichen Medien versuchen, Shell als ökologisch denkendes Positivbeispiel darzustellen (vgl. Monbiot, 2009). Gleichzeitig sind die Hauptprodukte von Shell umweltschädlich und die Investitionen in alternative Energien machen nur einen minimalen Prozentsatz der gesamten Unternehmensinvestitionen in die Produktion von Erdöl aus (vgl. Monbiot, 2009). In diesem Fall werden bewusst nur eine begrenzte Auswahl von Informationen dargestellt, um die Aufmerksamkeit der Konsumenten von anderen bedeutenden negativen Umweltauswirkungen abzulenken.

Institutionelle Interessenvertreter, die sich mit irreführender Werbung befassen, darunter die Europäische Kommission (Richtlinie 2005/29/EG über unlautere Geschäftspraktiken) und die US Federal Trade Commission (FTC) schreiben eigentlich vor, dass Unternehmen eine klare und deutliche Sprache verwenden sollen, um zu vermitteln, dass sich eine allgemeine Umweltaussage nur auf einen spezifischen und begrenzten Umweltnutzen bezieht (vgl. Federal Trade Commission, 2012, S. 62122). Die Regulierungsversuche solcher Organisationen sind jedoch von Land zu Land sehr unterschiedlich und ihre Durchsetzung ist im Allgemeinen ungenügend (vgl. Delmas & Burbano, 2011, S. 70-72). Aufgrund dessen argumentieren Wissenschaftler und Umweltschützer, dass solche unverbindlichen Regulierungsrichtlinien Verbraucher nur unzureichend vor den nachteiligen Auswirkungen des Greenwashing schützen (vgl. Feinstein, 2013, S. 229).

Konsumenten sollten sich in der Regel bei ihren Kaufentscheidungen auf Werbeaussagen und Botschaften von Unternehmen verlassen können, doch Greenwashing untergräbt dieses Vertrauen in die Werbemaßnahmen (vgl. Hamann & Kapelus, 2004, S. 86). Wenn die Konsumenten den Behauptungen von Unternehmen nicht mehr vertrauen können, sind sie auch nicht in der Lage, umweltfreundlich einzukaufen, da sie nicht wissen, welche Unternehmen und welche Produkte tatsächlich umweltfreundlich sind. Infolgedessen gefährdet Greenwashing schon jetzt den gesamten Markt für nachhaltige Produkte und Dienstleistungen und könnte den tatsächlich umweltfreundlichen Unternehmen langfristig erheblichen Schaden zufügen (vgl. Y.-S. Chen & Chang, 2013, S. 489).

2.3.1. Motive und Treiber von Greenwashing

Die ansteigende Verbreitung von Greenwashing kann also tiefgreifende negative Auswirkungen auf das Vertrauen der Konsumenten in umweltfreundliche Produkte haben und den Markt für umweltfreundliche Produkte und Dienstleistungen stark beanspruchen (vgl. Furlow, 2010, S. 23-24). Weiter kann es das Vertrauen der Anleger in umweltfreundliche Unternehmen negativ beeinflussen und den sozial verantwortlichen Kapitalmarkt für Investitionen untergraben. Greenwashing birgt auch einige Risiken, wenn Verbraucher, Nonprofit-Organisationen oder staatliche Stellen die Umweltaussagen von Unternehmen in Frage stellen. Zum Beispiel wurde die Green Mountain Power Corporation von mehreren Umweltgruppen ins Visier genommen, weil sie offenbar umweltverschmutzende Verbrennungstechnologien für ihre erneuerbaren Energiequellen, die sie als „grüne Energie“ vermarkteten, eingesetzt hat (vgl. Delmas & Burbano, 2011, S. 65). Ebenso wurden Unternehmen verklagt, weil sie sich an ökologischer Werbung beteiligt hatten, ohne berechtigte Gründe dafür zu haben. Bspw. hat Honda eine Sammelklage wegen falscher und irreführender Aussagen über die Kraftstoffeffizienz eines Hybridfahrzeugs beigelegt (vgl. Lane, 2010, S. 752-757). Greenwashing hat folglich zahlreiche potenzielle negative Konsequenzen für Unternehmen und den gesamten Markt für nachhaltige Produkte. Daher muss die Frage gestellt werden, warum trotz dieser Risiken so viele Unternehmen Greenwashing anwenden.

Um dies zu beantworten muss zunächst eine kurze Typologie erfolgen. Unternehmen, die positiv über ihre Umweltleistung kommunizieren, zum Beispiel durch Marketing- und PR-Kampagnen, werden als „vokale“ Unternehmen bezeichnet, während diejenigen, die nicht über ihre Umweltleistung kommunizieren, „stille“ Unternehmen genannt werden. Weiter sind Unternehmen, die eine gute Umweltleistung haben und diese auch kommunizieren „vokale grüne Unternehmen“ (Quadrant II), während diejenigen, die nicht über ihre Umweltleistung kommunizieren, als „stille grüne Unternehmen“ (Quadrant IV) bezeichnet werden können. Unternehmen mit einer schlechten Umweltleistung werden „braune Unternehmen“ genannt. Hier werden diejenigen, die nicht über ihre Umweltleistung kommunizieren, als „stille braune Unternehmen“ (Quadrant III) bezeichnet. Braune Unternehmen, die positiv über ihre Umweltleistung kommunizieren, sind nun genau die, die in dieser Diskussion von Interesse sind. Dies sind nämlich Unternehmen, die Greenwashing betreiben (Quadrant I) (vgl. Delmas & Burbano, 2011, S. 67-68).

Greenwashing wird hier demnach als die Schnittmenge von zwei Unternehmensverhaltensweisen definiert: Schlechte Umweltleistung und positive Kommunikation über die Umweltleistung. Da die Treiber einer CSR-Strategie, also strategische sowie ökologische bzw. soziale Motive bekannt sind, liegt der Fokus auf der Kommunikation über die Umweltleistung von Unternehmen. Das heißt, es werden die Treiber beschrieben, welche Unternehmen mit schlechter Umweltleistung („braune“ Unternehmen) dazu bringen, positiv über ihre Umweltleistung zu kommunizieren.

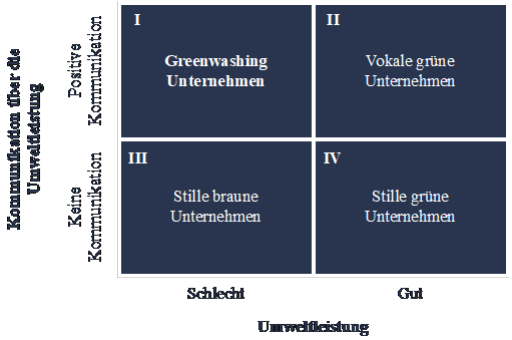


Abbildung 3: Typologie von Unternehmen basierend auf Umweltleistung und Kommunikation

Quelle: Eigene Darstellung in Anlehnung an Delmas & Burbano, 2011, S. 67

Delmas und Burbano (2011, S. 68) haben dazu ein vielzitiertes Modell entworfen, welches verschiedene Kategorien von Treibern unterscheidet (Abb. 4). Die erste Kategorie sind nicht-marktbezogene externe Treiber. Eine obligatorische Offenlegung von Umweltpraktiken und die Überprüfung solcher Informationen durch Dritte würde ein erfolgreiches Greenwashing erheblich erschweren. Selbst wenn die Greenwashing-Praktiken selbst nicht reglementiert wären, könnten Verbraucher, Investoren und Nonprofit-Organisationen die Kommunikation eines Unternehmens mit zuverlässigen Informationen über die Umweltpraktiken des Unternehmens vergleichen. Der derzeitige Stand der sehr limitierten Regulierung und der freiwilligen Offenlegung von Umweltinformationen durch die Unternehmen schreckt jedoch wenig von Greenwashing ab, weshalb die Versuchung groß ist. Aktivisten, Nonprofit-Organisationen und die Medien drohen mit öffentlicher Bloßstellung des Greenwashing, was wahrscheinlich einige braune Firmen davon abhält, positiv über ihre Umweltleistung zu kommunizieren. Fakt ist aber, dass der drohende Imageverlust viele Unternehmen nicht ausreichend abschreckt.

Die nächste Kategorie sind marktbezogene externe Treiber. Dies sind Verbraucher-, Investoren- und wettbewerbsinduzierte Anreize. Immer mehr Konsumenten und Kapitalgeber legen Wert auf Nachhaltigkeit, wodurch die Stellung von Unternehmen ohne gute Umweltleistung am Absatz- und Kapitalmarkt immer schwieriger wird. Je größer der Druck für „braune Firmen“ wird, umweltfreundlich zu wirken, desto größer werden die Anreize, positiv über ihre Umweltleistung zu kommunizieren (vgl. Vos, 2009, S. 682-683).

Darüber hinaus ist auch das Wettbewerbsumfeld ein kritischer Teil des Marktumfelds, in dem ein braunes Unternehmen vor der Entscheidung steht, ob es positiv über seine Umweltleistung kommunizieren will. Unternehmen neigen dazu, sich an ähnlichen Unternehmen in ihrer Branche zu orientieren, die sie für legitimer oder erfolgreicher halten und Untersuchungen haben gezeigt, dass dies auch in Bezug auf die Umweltleistung zutrifft (vgl. Delmas & Toffel, 2008, S. 1034-1035). Dies deutet darauf hin, dass einige Firmen über vermeintlich umweltfreundliche Praktiken kommunizieren,

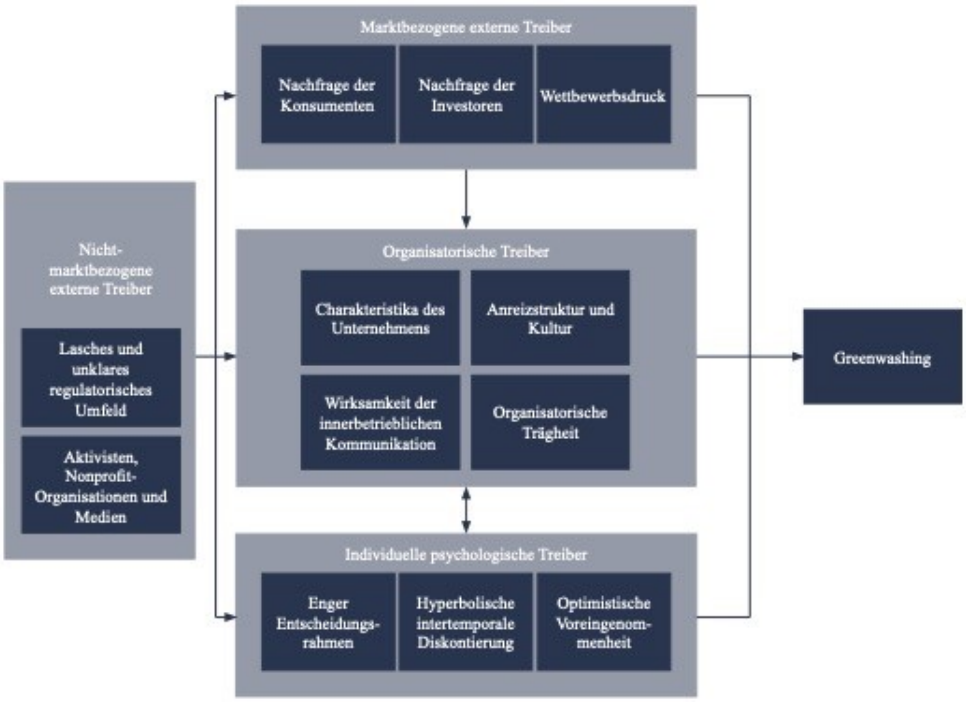


Abbildung 4: Die Treiber von Greenwashing

Quelle: Eigene Darstellung in Anlehnung an Delmas & Burbano, 2011, S. 68

aus Angst, den Anschluss an ihre Konkurrenten zu verlieren, die bereits positiv über ihre Umweltleistung kommunizieren.

Die dritte Kategorie sind die Treiber auf Organisationsebene. Diese sind die Charakteristika des Unternehmens, die Anreizstruktur und Kultur, die Wirksamkeit der innerbetrieblichen Kommunikation und die organisatorische Trägheit.

Merkmale auf Unternehmensebene (z.B. Größe, Branche, Rentabilität) beeinflussen zweifellos die Gesamtstrategien. Konsumgüterunternehmen sehen sich wahrscheinlich einem größeren Druck seitens der Verbraucher ausgesetzt, umweltfreundlich zu erscheinen, als andere. Ebenso erzeugen große, finanzstarke Firmen eher gesellschaftliches Interesse und stehen daher unter einem größeren Druck von Seiten der Konsumenten und Investoren als kleinere Unternehmen (vgl. Delmas & Burbano, 2011, S. 73).

Darüber hinaus hat sich gezeigt, dass eine feste Anreizstruktur und ein ethisches Klima Determinanten eines aufrichtigen Verhaltens sein können (vgl. Wimbush, Shepard & Markham, 1997, S. 1714-1715). Unethisches Verhalten tritt häufiger in Unternehmen auf, in denen egoistische im Gegensatz zu prinzipientreuen ethischen Klimata vorherrschen. Zur Verbesserung können Ethikkodizes und andere explizite Verhaltensstandards in der Unternehmensstruktur implementiert werden (vgl. Cullen, Parboteeah & Victor, 2003, S. 127).

Auch ein unzureichender Wissenstransfer innerhalb einer Firma kann zu (unbeabsichtigtem) Greenwashing führen (vgl. Szulanski, 1996, S. 27). Beispielsweise könnte eine Marketingabteilung die Umweltfreundlichkeit eines Produkts aufgrund mangelnder Kommunikation mit einer Produktentwicklungs-, oder Verpackungsabteilung überbewerten.

Weiterhin hat die Managementliteratur zunehmend die Trägheit von Organisationen als einen Faktor erkannt, der das Verhalten von Unternehmen beeinflusst (vgl. Rumelt, 1995, S. 102). So könnte sich bspw. Bob Dudley, der ehemalige Vorstandsvorsitzende von BP, zum Teil aufgrund organisatorischer Trägheit an einer Form von Greenwashing beteiligt haben. Er ist von vielen Medien kritisiert worden, weil er ausgiebig über die Verbesserung der Sicherheit sprach, nachdem er nach der Explosion eines Bohrturms (Deepwater Horizon) Vorstandsvorsitzender wurde, aber bis dahin seinen Worten kaum Taten folgen ließ. Ein Grund dafür war, dass trotz seiner Absicht, Prozesse zu ändern, die Umsetzung solcher Änderungen aufgrund organisatorischer Trägheit länger dauerte als erwartet (vgl. Delmas & Burbano, 2011, S. 74).

Außerdem existieren psychologische Treiber auf individueller Ebene. Führungskräfte und Einzelpersonen spielen eine wichtige Rolle bei der Erklärung von Firmenverhalten. Veranlagungen wie ein enger Entscheidungsrahmen, hyperbolische intertemporale Diskontierung und optimistische Voreingenommenheit treten immer stärker hervor und haben einen großen Einfluss auf die individuelle Entscheidungsfindung. Folge ist eine so genannte eingeschränkte Rationalität (vgl. Kahneman, 2003, S.1449).

Ein enger Entscheidungsrahmen beschreibt die Neigung, Entscheidungen in Isolation zu treffen (vgl. Kahneman & Lovallo, 1993, S. 19). So könnte eine Führungskraft eines Un-

ternehmens heute beschließen, die Umweltfreundlichkeit eines Produkts positiv zu kommunizieren, ohne angemessen zu berücksichtigen, was erforderlich ist, um dies umzusetzen - was zwangsweise zu Greenwashing führt.

Hyperbolische intertemporale Diskontierung beschreibt kurz gesagt die aus der VWL bekannte Gegenwartspräferenz (vgl. Rubinstein, 2003, S. 1208). Beispielsweise möchte ein Manager nicht warten, bis das Unternehmen tatsächlich ökologisch handelt und veranlasst daher schon vorzeitig eine positive Kommunikation für eine Umweltleistung, die ggf. nie eintritt.

Optimistische Voreingenommenheit, d.h. die Neigung von Einzelpersonen, die Wahrscheinlichkeit positiver Ereignisse zu überschätzen und die Wahrscheinlichkeit negativer Ereignisse zu unterschätzen, kann ebenfalls zum Greenwashing beitragen (vgl. Helweg-Larsen & Shepperd, 2001, S. 74). Entscheidungsträger können die Wahrscheinlichkeit der positiven Ergebnisse des Greenwashing, wie die Chance, neues Kapital von umweltbewussten Investoren anzuziehen, überschätzen, und die Wahrscheinlichkeit negativer Ereignisse infolge von Greenwashing unterschätzen.

Zusammengefasst kann es also sehr unterschiedliche Gründe haben, weshalb Unternehmen trotz der Risiken Greenwashing praktizieren. Aktuell sorgen wohl insbesondere marktbezogene externe Treiber dafür, dass dieses Thema so bedeutend ist wie nie zuvor. Die zunehmende Nachfrage nach nachhaltigen Produkten und Dienstleistungen sowohl von Konsumenten als auch von Investoren verleitet viele Unternehmen dazu, falsche Informationen zur eigenen Umweltleistung zu vermitteln (vgl. Parguel et al., 2011, S. 15).

2.4. Unternehmensreputation

Die Reputation von Unternehmen ist in jüngster Zeit zu einer der bedeutendsten strategischen Komponenten für den Erfolg auf wettbewerbsorientierten Märkten geworden. Eine positive Reputation ist zweifelsohne eine zentrale Zielgröße der Unternehmenskommunikation und ein wesentlicher Faktor für das langfristige Bestehen eines Unternehmens am Markt (vgl. Keh & Xie, 2009, S. 732; K.-H. Kim, Kim & Qian, 2018, S. 1112; Raithel & Schwaiger, 2015, S. 945).

Die von den Stakeholdern wahrgenommene Unternehmensreputation ist deshalb so wichtig, weil sie sowohl kommerzielle Größen wie bspw. Umsatz oder Gewinn, als auch nicht-kommerzielle Größen, wie Markenvertrauen oder -loyalität beeinflusst (vgl. G. Walsh & Wiedmann, 2004, S. 304). Außerdem hat eine gute Reputation einen positiven Effekt auf den Marktwert von Unternehmen. Investoren sind eher bereit, ihre Einlagen Firmen anzuvertrauen, die aufgrund geringerer wahrgenommener Risiken und einer potenziell wirksameren Kommunikation einen besseren Ruf genießen. Auch die Marketinganstrengungen des Unternehmens können erheblich verbessert werden, da qualitätsempfindliche Segmente mit weniger Preisverhandlungen und potenziell niedrigeren Verkaufskosten angesprochen werden können (vgl. C. Fombrun & Shanley, 1990, S. 252-253; Miles & Covin, 2000, S. 300). Es ist also klar erkennbar, warum es für Unternehmen so wichtig ist, zu verstehen, was eine gute Reputation ausmacht und wie diese erreicht werden kann.

Definiert wird Unternehmensreputation dabei als eine kollektive Darstellung des bisherigen Verhaltens und der Resultate eines Unternehmens, die die Fähigkeit des Unternehmens widerspiegelt, mehreren Interessengruppen wertvolle Ergebnisse zu liefern (vgl. C. J. Fombrun & Rindova, 2001, S. 7). Sie misst die relative Stellung eines Unternehmens intern bei den Mitarbeitern und extern bei den weiteren Stakeholdern, sowohl im internen als auch im Wettbewerbsumfeld (vgl. C. J. Fombrun, Gardberg & Sever, 2000, S. 243). Dabei geht sie insofern über das Unternehmensimage hinaus, als dass die Reputation nicht nur das aktuelle Bild umfasst, welches die Stakeholder von einem Unternehmen haben, sondern auch die in der Vergangenheit aufgebauten und für die Zukunft relevanten Unterstützungspotenziale, wie bspw. Vertrauen, Interesse oder Zutrauen (vgl. Wiedmann et al., 2006, S. 99). Der Begriff Unternehmensreputation umfasst also alle unternehmensbezogenen Wahrnehmungen, Einschätzungen und daraus hervorgehenden Einstellungen aller Stakeholder (vgl. Wiedmann, 2012, S. 59).

Das Interesse von Wissenschaft und Praxis an der Reputation von Unternehmen lässt sich auch auf die weite Verbreitung des Themas durch die regelmäßige Veröffentlichung von Unternehmensratings zurückführen. Fortunes „Americas Most Admired Companies“, die Financial Times „World’s Most Respected Companies“ oder Studien des Manager Magazins bewerten Unternehmen hinsichtlich ihrer Reputation. Leider kommt es hierbei zu inkonsistenten Ergebnissen, die auf methodische Mängel sowie Messprobleme zurückzuführen sind (vgl. C. J. Fombrun et al., 2000, S. 243-245). Grund dafür ist, dass diese Ratings in erster Linie auf den Wahrnehmungen von Führungskräften und Finanzanalysten beruhen, die in der Regel gebeten werden, ein Unternehmen nach einer begrenzten Anzahl von Kriterien zu bewerten. Um diese Limitationen aufzuheben benötigt es ein neues, robusteres Maß für die Unternehmensreputation, das sowohl valide als auch reliabel ist (vgl. Dowling & Gardberg, 2012, S. 34-35).

Daher wurde zunächst der Reputationsquotient (RQ) entwickelt, eine sechsdimensionale Skala, die aus 20 Attributen konstruiert wurde (vgl. C. J. Fombrun et al., 2000, S. 253). Aus diesem wurde das vier Attribute umfassende RepTrak Pulse-Maß ausgegliedert und verwendet, um ein separates Maß für die emotionale Bindung einer Person an ein Unternehmen zu schaffen (vgl. Ponzi, Fombrun & Gardberg, 2011, S. 16; Sarstedt, Wilczynski & Melewar, 2013, S. 329).

Das vollständige RepTrak-Konzept wurde 2005-2006 entwickelt, um Führungskräften ein Analyseinstrument bereitzustellen, mit dem nicht nur die Wahrnehmung eines Unternehmens durch Stakeholder verfolgt und bewertet werden kann, sondern welches auch ein umfassenderes Verständnis der zugrundeliegenden informationellen Treiber der Reputation ermöglicht, die eine emotionale Bindung hervorrufen. Der RepTrak besteht aus den sieben Dimensionen (1) Produkte; (2) Innovation; (3) Arbeitsumfeld; (4) Leitung; (5) Gesellschaftliche Einbindung; (6) Führung; (7) Performance (vgl. C. J. Fombrun, Ponzi & Newburry, 2015, S. 4-8) und

wird in dieser Studie als Messinstrument für die Effekte unterschiedlicher Werbeanzeigen verwendet.

3. Theoretische Ansätze

3.1. Die Effekte von CSR auf die Unternehmensreputation

Zunächst ist es wichtig, bei der Verwendung von CSR zur Differenzierung von Produkten zwischen zwei Arten zu unterscheiden. Die erste ist die vertikale Differenzierung. Diese tritt dann auf, wenn der Großteil der Konsumenten ein Produkt einem anderen vorziehen. Bspw. bevorzugen die meisten Verbraucher bei ansonsten gleichen Voraussetzungen ein kraftstoffeffizienteres Fahrzeug. Im Zusammenhang mit CSR könnte eine solche Situation auftreten, wenn in der Denkweise der Konsumenten klar ist, dass das Produkt mit einem CSR-Merkmal besser ist als das Produkt ohne ein solches Merkmal. Eine „Hybrid"-Version eines Audi A6 erzeugt bspw. weniger Umweltverschmutzung als ein standardmäßiger Audi A6. Somit ist für die meisten Verbraucher klar, dass das Hybridauto besser ist als das Standardmodell. Einige Verbraucher sind bereit, einen Preisaufschlag für das Hybridauto zu zahlen, da die soziale Eigenschaft der geringeren Umweltverschmutzung wertvoll für sie ist. Diese Art der Differenzierung kann die Reputation eines Unternehmens stärken, was einen Mehrwert schafft und es dem Unternehmen zusätzlich ermöglicht, eine bestimmte Marktnachfrage zu befriedigen (vgl. C. Fombrun & Shanley, 1990, S. 233).

Im Gegensatz dazu findet eine horizontale Differenzierung statt, wenn nur einige wenige Konsumenten ein bestimmtes Produkt bevorzugen, die Entscheidung jedoch auf den Geschmack und nicht auf die Qualität zurückzuführen ist. Beispielsweise entscheiden sich einige Konsumenten aufgrund der Farbe für ein bestimmtes Fahrzeug. Diese Art der Differenzierung trägt nicht zur Unternehmensreputation bei und ermöglicht es dem Unternehmen nicht, einen höheren Preis zu verlangen. Die horizontale Differenzierung gilt auch für verschiedene Marken. Beispielsweise bevorzugen einige Verbraucher Coca-Cola gegenüber Pepsi, während andere die gegenteilige Ansicht vertreten. Die Gründe dafür liegen ebenfalls nicht bei der Qualität der Produkte, sondern dem subjektiven Empfinden der Konsumenten (vgl. McWilliams et al., 2006, S. 4-5).

In dieser Studie wird demnach die vertikale Produktdifferenzierung betrachtet. Die in dem Experiment dieser Studie verwendeten Produkte unterscheiden sich ausschließlich in ihrem gesellschaftlichen Mehrwert, alle anderen Attribute sind identisch.

Weiterhin muss zwischen überzeugender und informativer CSR-Werbung zu unterschieden werden. Überzeugende CSR-Werbung versucht, den Geschmack der Verbraucher für Produkte mit CSR-Attributen positiv zu beeinflussen. Daraus folgt, dass diese Art von Werbung nicht firmenspezifisch sein muss. Informative CSR-Werbung hingegen gibt gezielt Auskunft über die CSR-Merkmale oder die CSR-Praktiken eines Unternehmens (vgl. McWilliams et al., 2006, S. 5). Die vorliegende Studie verwendet informative CSR-Werbung, da die jeweiligen Effekte auf die Unternehmensreputation gemessen und der Konsument nicht allgemein von CSR überzeugt werden soll.

CSR ist bereits seit einigen Jahren zu einer bedeutenden gesellschaftlichen Verpflichtung in diversen Geschäftsbereichen geworden, weshalb in einer Reihe von früheren Studien die Einflüsse von CSR auf die Unternehmensleistung untersucht worden sind (vgl. Wang, Dou & Jia, 2016, S. 1097-1101). Aus der Perspektive des Managements sind CSR-Aktivitäten eine wesentliche Aufgabe für das nachhaltige und strategische Überleben von Unternehmen (vgl. Kolk, 2016, S. 23). Um zu erklären, wie CSR die Reputation einer Organisation beeinflusst, kann die Signaltheorie betrachtet werden (vgl. Walker, 2010, S. 376). Die Signaltheorie ist hilfreich, um das Verhalten zu beschreiben, wenn zwei Parteien (Individuen oder Organisationen) Zugang zu asymmetrischen Informationen haben. Typischerweise muss eine Partei, der Sender, wählen, ob und wie er diese Informationen kommuniziert (oder signalisiert), und die andere Partei, der Empfänger, muss wählen, wie er das Signal interpretiert. Dabei sorgen unterschiedliche Kommunikationsmethoden in der Regel für unterschiedliche Interpretationen und Reaktionen bei dem Empfänger des Signals (vgl. Connelly, Certo, Ireland & Reutzel, 2011, S. 39).

Bspw. besetzen die Leiter einer jungen Firma, die sich in einem Börsengang befindet, ihren Vorstand mit einer vielfältigen Gruppe angesehener Geschäftsführer, um potenziellen Investoren eine Botschaft über die Legitimität der Firma zu vermitteln (vgl. Certo, 2003, S. 432; Filatotchev & Bishop, 2002, S. 941). Unterschiedliche Signale können dabei ganz verschiedene Reaktionen hervorrufen. Daraus ergibt sich für diese Studie folgende erste Hypothese:

> *H_1: Unterschiedliche CSR-Initiativen haben unterschiedliche Effekte auf die Unternehmensreputation.*

Im Zusammenhang mit CSR konnte Folgendes festgestellt werden: Konsumenten suchen, wenn sie mit einer Informationsasymmetrie konfrontiert sind, nach Signalen, welche verantwortungsbewusste von unverantwortlichen Unternehmen trennen. Verschiedene CSR-Initiativen eines Unternehmens erzeugen demnach positive Signale, welche die Glaubwürdigkeit und Reputation verbessern können (vgl. Basdeo, Smith, Grimm, Rindova & Derfus, 2006, S. 1216; Pfau, Haigh, Sims & Wigley, 2008, S. 150).

A. A. King, Lenox und Terlaak (2005, S. 1102) erklären bspw., dass die Zertifizierung privater Managementnormen, wie die Zertifizierung des Umweltmanagementsystems nach ISO 14001 ein Signal sendet, welches die nicht beobachtbaren Merkmale eines Unternehmens übermittelt. Dieses Signal bildet glaubwürdige Informationen und verringert dadurch Informationsasymmetrien. Auch eine umfangreiche Studie von Mishra und Suar (2010, S. 585) kommt zu dem Ergebnis, dass verantwortungsvolle Geschäftspraktiken gegenüber Stakeholdern für Unternehmen profitabel und vorteilhaft sein können. Es gibt endlos viele Möglichkeiten an

CSR-Initiativen, die Unternehmen durchführen können. Im heutigen hart umkämpften Marktumfeld haben zahlreiche Unternehmen CSR als strategisches Instrument erkannt, um auf die Erwartungen verschiedener Stakeholder wie Medien, Nonprofit-Organisationen und Konsumenten zu reagieren und so ein vorteilhaftes Unternehmensimage zu schaffen (vgl. Jones, 2005, S. 17-19). Grundsätzlich zeigt sich dabei, dass alle CSR-Initiativen das Potential haben, die Unternehmensreputation zu verbessern (vgl. Saeidi, Sofian, Saeidi, Saeidi & Saaeidi, 2015, S. 347-348; Miles & Covin, 2000, S. 308-309).

Darauf aufbauend zeigen andere empirische Studien, dass eine gute Reputation die Kaufbereitschaft von Konsumenten fördert (vgl. Brown & Dacin, 1997, S. 68; Alexander, 2002, S. 234). Während sich eine schlechte Reputation letztlich negativ auf die Gesamtproduktbewertung auswirkt, kann eine positive Reputation die Produktbewertung verbessern (vgl. Brown & Dacin, 1997, S. 68).

Es wird also davon ausgegangen, dass, wenn sich ein Unternehmen um die unterschiedlichen Interessengruppen bemüht, diese positiv reagieren und das Unternehmen selbst positiv bewerten (vgl. X. Chen & Kelly, 2015, S. 112) wodurch sich sowohl Faktoren wie Image, Loyalität und Markenidentifikation (vgl. Javed, Rashid, Hussain & Ali, 2020, S. 1403) als auch die Unternehmensreputation und die Kaufabsicht der Konsumenten erhöhen kann (vgl. Aguinis & Glavas, 2012, S. 959-960; S. Du, Bhattacharya & Sen, 2010, S. 17-18; V. Smith & Langford, 2009, S. 107; Torres, Bijmolt, Tribó & Verhoef, 2012, S. 22-23). Aufgrund dieser Ergebnisse aus unterschiedlichen Studien kann folgende Hypothese geschlussfolgert werden:

> *H_2: CSR-Initiativen haben insgesamt positive Effekte auf die Unternehmensreputation.*

In den letzten Jahrzehnten scheint es aufgrund der wachsenden Sensibilität in Bezug auf Umweltbelange auch zu einer wachsenden Skepsis der Stakeholder gegenüber Unternehmen, die ihre Umweltstrategien und -leistungen kommunizieren, gekommen zu sein (vgl. Do Paço & Reis, 2012, S. 147; Bickart & Ruth, 2012, S. 52; Royne, Martinez, Oakley & Fox, 2012, S. 95).

Skepsis ist eine mögliche kognitive Reaktion auf Werbeeindrücke (vgl. Pomering & Johnson, 2009, S. 427). Kognitive Reaktionen sind nachrichtenrelevante Gedanken, die während einer Überlegung entstehen, wenn man Nachrichteninhalte mit Vorwissen und im Gedächtnis gespeicherten Haltungen in Beziehung setzt (vgl. Meyers-Levy & Malaviya, 1999, S. 47). In der wissenschaftlichen Literatur ist dieses Verhalten als „Perceived Consumer Scepticism“ (PCS), also die wahrgenommene Skepsis der Konsumenten, bekannt. Sie wird allgemein definiert als die Skepsis der Konsumenten gegenüber umweltbezogener Werbung aufgrund der Häufigkeit von irreführenden Aussagen, also Greenwashing (vgl. Matthes & Wonneberger, 2014, S. 116). Konsumenten sind gegenüber Werbung skeptischer als gegenüber jeder anderen Form der Kommunikation (vgl. Obermiller, Spangenberg & MacLachlan, 2005, S. 7). Das Konstrukt des PCS, welches von Mohr, Erolu und Ellen (1998, S. 46) eingeführt wurde, bezieht sich auf die Neigung der Verbraucher, Umweltaussagen in der Werbung nicht länger zu vertrauen (vgl. Matthes & Wonneberger, 2014, S. 116). Mehrere Studien bestätigen, dass Konsumenten misstrauisch gegenüber dem Wahrheitsgehalt von Umweltaussagen von Unternehmen sind und über die hintergründigen Absichten spekulieren (vgl. Foreh & Grier, 2003, S. 354-355; Yoon, Gürhan-Canli & Schwarz, 2006, S. 388; De Vries, Terwel, Ellemers & Daamen, 2015, S. 151; Anuar, Omar & Mohamad, 2013, S. 96-98). Sehr problematisch ist, dass Greenwashing die Skepsis so sehr erhöht hat, dass es vorkommen kann, dass selbst tatsächliche CSR-Initiativen von Unternehmen für eine Form von Greenwashing gehalten werden (vgl. Jahdi & Acikdilli, 2009, S. 111). Oft haben Verbraucher nicht das Fachwissen oder die Fähigkeit, die Umwelt- und Verbraucherwerte scheinbar umweltfreundlicher Produkte zu überprüfen, was zu Fehlwahrnehmungen und Skepsis führt (vgl. Ottman, Stafford & Hartman, 2006, S. 31). Um dem möglichst entgegenzuwirken werden daher von Unternehmensseite häufig Informationen bereitgestellt, welche die tatsächlich nachhaltige Unternehmensleistung beweisen sollen. Dennoch ist die Skepsis der Konsumenten ein großes Problem im Green Marketing (vgl. Nyilasy, Gangadharbatla & Paladino, 2014, S. 693; Skarmeas & Leonidou, 2013, S. 1849-1850). Die möglichen Einflüsse von Konsumentenskepsis auf die Wirksamkeit von CSR-Initiativen müssen daher auch in dieser Studie beachtet werden, woraus sich die Hypothese ergibt:

> *H_3: CSR-Initiativen haben negative Effekte auf die Unternehmensreputation, wenn sie fälschlicherweise für Greenwashing gehalten werden.*

Außerdem spielt auch die wahrgenommene Motivation der Unternehmen eine Rolle. Die strategischen Motivatoren für CSR-Initiativen wurden bereits ausführlich beschrieben, jedoch kann und sollte CSR aber auch immer intrinsisch motiviert sein. Dass ein Unternehmen in der Tat eine moralische Verpflichtung gegenüber der Gesellschaft hat, sich verantwortungsvoll zu verhalten, kann durch verschiedene ethische Theorien wie die kantische Ethik (vgl. Evan & Freeman, 1988, S. 75; Bowie, 2017) oder die Tugendethik (vgl. Solomon, 1992) begründet werden. Chandler (2016, S. 248) erklärt weiter, dass strategisches CSR nur dann effizient ist, wenn es so in die strategische Planung eines Unternehmens eingegliedert wird, dass das Unternehmen im Interesse einer breiten Palette von Stakeholdern geführt wird und dadurch langfristige Wertoptimierung anstrebt. Demnach werden CSR-Initiativen negativer bewertet, wenn sie als ausschließlich strategisch und egoistisch wahrgenommen werden (vgl. Graafland & Van de Ven, 2006, S. 121). Die dabei gebildeten Attributionen können sich wiederum negativ auf die Einstellung zur Marke und die Kaufabsicht auswirken (vgl. Walker, 2010, S. 672). Die Reaktionen der Konsumenten auf eine CSR-Initiative können also stark davon abhängen, welche Motivation dem Unternehmen zugeschrieben wird.

Extrinsische Attributionen fungieren dabei wahrscheinlich als negative Moderatoren für die Beziehung zwischen CSR-Wahrnehmung und CSR-Glaubwürdigkeit (vgl. S. Du, Bhattacharya & Sen, 2007, S. 238). Asymmetrischere Informationen machen es allerdings schwierig, die Motivation der CSR-Initiativen zu erkennen. So werden Unternehmen versuchen, ihre strategischen Motivationen zu verschleiern, um die Effektivität der CSR-Initiativen zu erhöhen (vgl. McWilliams et al., 2006, S. 9). Für diese Studie ergibt sich daraus die Hypothese:

> *H_4: Als intrinsisch motiviert wahrgenommene CSR-Initiativen haben stärkere positive Effekte auf die Unternehmensreputation als solche, die als extrinsisch motiviert wahrgenommen werden.*

Frühere Studien (vgl. McQuarrie & Mick, 2003, S. 218; A. A. Mitchell, 1986, S. 16-21; Burns, Biswas & Babin, 1993, S. 71) haben weiterhin gezeigt, dass visuelle Kommunikationsmethoden mehr Aufmerksamkeit erregen, reichere Schlussfolgerungen hervorrufen und angenehmer und überzeugender sind als schriftliche Aussagen. Aufgrund dessen kann davon ausgegangen werden, dass Umweltlabel grundsätzlich eine stärkere Reaktion bei den Konsumenten hervorrufen als rein schriftliche CSR-Initiativen. Als weitere Hypothese ergibt sich daher:

> *H_5: CSR-Initiativen mit einem Umweltlabel haben stärkere positive Effekte auf die Unternehmensreputation als CSR-Initiativen ohne Umweltlabel.*

3.2. Die Effekte unterschiedlicher Greenwashing-Methoden auf die Unternehmensreputation

Um die Effekte unterschiedlicher Greenwashing-Methoden auf die Unternehmensreputation untersuchen zu können, müssen diese Methoden zunächst identifiziert und definiert werden. Greenwashing wird durch Kommunikationsmedien ausgeführt; in erster Linie durch Werbeanzeigen. In der Literatur werden zwei Arten von Greenwashing-Werbung definiert: Claim-Greenwashing und Executional-Greenwashing. Claim-Greenwashing erfolgt durch die Verwendung unbestimmter oder strittiger Begriffe, die Erzeugung falscher Behauptungen oder die Manipulation von Behauptungen, um notwendige Informationen auszuschließen (vgl. Kangun, Carlson & Grove, 1991, S. 48). Mit anderen Worten, Claim-Greenwashing ist „Lügen, Lügen durch Weglassen oder Lügen durch Unklarheit“ (Parguel et al., 2015, S. 108). Auf der anderen Seite geht es beim Executional Greenwashing um die Verwendung von naturanregenden Elementen in der Werbung, wie bspw. Naturbildern unter Verwendung von Farben (z.B. blau, grün), Klängen (z.B. Vögel, das Meer) und natürlichen Landschaften (z.B. Berge, Wälder) (vgl. Hartmann & Apaolaza-Ibáñez, 2009, S. 715; Parguel et al., 2015, S. 2).

In dieser Studie liegt der Fokus auf dem Claim-Greenwashing. Mithilfe einer umfangreichen Studie in Kanada und den USA über mehrere Jahre hat TerraChoice sieben verschiedene Methoden von Claim-Greenwashing definiert und diese als „Die 7 Sünden des Greenwashing“ publik gemacht. TerraChoice ist eine in Kanada ansässige Agentur für Umweltmarketing. Sie gehört zur Underwriter Laboratories Group, einer der größten unabhängigen Test- und Zertifizierungsorganisationen der Welt (vgl. TerraChoice, 2010, S. 3). Diese sieben Sünden sind:

1. Die Sünde des versteckten Kompromisses
   - wird begangen, wenn ein Unternehmen nur eine begrenzte Auswahl von Qualitäten darstellt, um die Aufmerksamkeit der Konsumenten von anderen bedeutenden negativen Umweltauswirkungen abzulenken. Zum Beispiel ist Papier nicht unbedingt umweltfreundlich, nur weil es aus einem nachhaltig bewirtschafteten Wald stammt. Andere Umweltprobleme im Papierherstellungsprozess wie Wasser- und Luftverschmutzung können mindestens genauso relevant sein.
2. Die Sünde der fehlenden Beweise
   - wird begangen, wenn ein Unternehmen Behauptungen aufstellt, die nicht durch leicht zugängliche Informationen oder durch eine verlässliche Zertifizierung durch Dritte belegt werden können.
3. Die Sünde der Vagheit
   - wird von Unternehmen begangen, die weit gefasste, irreführende Begriffe wie „rein“, „natürlich“ oder „organisch“ verwenden. Auch Uran oder Formaldehyd kommen ganz natürlich vor - und sind hochgiftig. „Natürlich“ bedeutet nicht zwangsläufig nachhaltig.
4. Die Sünde der Irrelevanz
   - wird begangen, wenn ein Unternehmen eine grüne Behauptung aufstellt, die entweder unbedeutend oder nicht hilfreich für den Konsumenten ist, der nachhaltige Produkte sucht. „FCKW-frei“ ist ein typisches Beispiel, da dies häufig beworben wird, obwohl FCKW seit 30 Jahren gesetzlich verboten ist.
5. Die Sünde des geringeren von zwei Übeln
   - wird vom Unternehmen begangen, wenn es eine Nachhaltigkeit bewirbt, die innerhalb der Produktkategorie zutrifft, die Produktkategorie an sich aber dennoch umweltschädigend ist. Bio-Zigaretten können ein Beispiel für diese Kategorie sein.
6. Die Sünde des Schwindels
   - eine vom Unternehmen begangene unwahre grüne Behauptung. Dies ist die seltenste Art von Greenwashing. Sie lässt sich insbesondere bei Umweltzertifizierungen wie dem „Energy Star“ finden, da hier keine Überprüfung erfolgt. Den „Energy Star“ kann jeder Hersteller verwenden,

der angibt, dass sein Gerät diesen Standard erfüllt. Gibt der Hersteller falsche Angaben an, um das Label verwenden zu können, so begeht er die Sünde des Schwindelns.

7. Die Sünde der Verwendung falscher Label
   - wird von Unternehmen begangen, die die Umweltfreundlichkeit des Produkts durch gefälschte oder erfundene Label und Zertifikate nachweisen, die den Anschein einer unabhängigen Überprüfung durch eine dritte Partei machen (vgl. TerraChoice, 2010, S. 10).

Zwei der am häufigsten verwendeten Methoden sind dabei die „Sünde der Vagheit“ sowie die „Sünde der Verwendung falscher Label“ (vgl. TerraChoice, 2010, S. 16). Aufgrund dessen werden diese in der vorliegenden Studie als unabhängige Variablen fungieren, um die Effekte unterschiedlicher Greenwashing-Methoden auf die Reputation von Unternehmen zu messen.

Für die theoretischen Ansätze dieser Effekte wird zunächst erneut die Signaltheorie betrachtet (vgl. Seele & Gatti, 2017, S. 239). Ausgehend von der Annahme, dass es die Intention von Unternehmen ist, effektive und effiziente Signale für die verschiedenen Adressaten zu schaffen, kann argumentiert werden, dass verschiedene Methoden des Greenwashing (bspw. die Verwendung falscher Label oder vager Formulierungen), die aus unterschiedlichen Ansätzen und strategischen Entscheidungen resultieren, unterschiedliche Ausmaße von Greenwashing produzieren. Demnach erzeugt jede Methode von Greenwashing, die als absichtlich irreführende Umweltkommunikation erkannt wird, unterschiedliche Reaktionen und Antworten von Stakeholdern (vgl. Collison et al. 2003, S. 208). So rufen einige Ansätze womöglich eher weniger gravierende Reaktionen hervor, während andere gegebenenfalls zu akuten Reputationsproblemen, Protesten oder Boykotten führen können (vgl. Torelli, Balluchi & Lazzini, 2020, S. 409).

Aufgrund dieser Erkenntnisse ergibt sich folgende Hypothese:

> *H_6: Unterschiedliche Greenwashing-Methoden haben unterschiedliche Effekte auf die Unternehmensreputation.*

Die empirische Forschung zu den Auswirkungen des Greenwashing ist noch begrenzt. Untersuchungen auf Makroebene legen allerdings nahe, dass Greenwashing keine positiven Auswirkungen auf die Gesamtleistungsindikatoren von Unternehmen hat. X. Du (2015, S. 547) beschreibt eine Analyse des chinesischen Aktienmarktes, aus der hervorgeht, dass Greenwashing eine negative Beziehung zu den kumulativen abnormalen Renditen (CAR) von Unternehmen hat, während CSR eine signifikant positive Beziehung zu CAR hat. Walker und Wan (2012, S. 227) haben die finanziellen Auswirkungen von Greenwashing für kanadische Firmen untersucht und dabei festgestellt, dass Greenwashing in einem negativen Zusammenhang mit der finanziellen Leistung steht. Zudem erklären Y.-S. Chen und Chang (2013, S. 497), dass Greenwashing ein geringeres „grünes Markenvertrauen“ zur Folge hat und Y.-S. Chen, Lin und Chang (2014, S. 2421) argumentieren, dass Greenwashing zu negativer Mundpropaganda führt. Zusätzlich dazu wirken sich wahrgenommene Täuschung und Skepsis seitens der Konsumenten auch negativ auf die Glaubwürdigkeit von Organisationen und die wahrgenommene Unternehmensleistung aus (vgl. Nyilasy et al., 2014, S. 693).

Wie Greenwashing die von den Konsumenten wahrgenommene Unternehmensreputation beeinflusst, ist in der Literatur nicht konkret beantwortet. Zusammengefasst deuten diese Studien aber darauf hin, dass Greenwashing-Methoden eher Skepsis bei den Konsumenten hervorrufen und keinen Mehrwert für Unternehmen hervorbringen. Es ist jedoch schwierig, aus solchen Daten kausale Schlussfolgerungen für dir Auswirkungen auf die Unternehmensreputation zu ziehen. Diese Ergebnisse aus der Literatur führen zu der Hypothese:

> *H_7: Greenwashing-Methoden haben insgesamt negative Effekte auf die Unternehmensreputation.*

Auf der anderen Seite haben De Jong, Harkink und Barth (2018) in ihrer Studie herausgefunden, dass Greenwashing tatsächlich das Potential hat, den Eindruck der Konsumenten von den Umweltaussagen und der Umweltleistung von Unternehmen positiv zu beeinflussen. Im Vergleich zu „stillen braunen Unternehmen“ bewerteten die Probanden Greenwashing-Unternehmen hinsichtlich ihrer Umweltleistung besser. Diese Bewertung ist zwar signifikant weniger positiv als die von „stillen grünen“ und „vokalen grünen Unternehmen“, aber dennoch zeigt sich, dass Greenwashing-Methoden auch positive Effekte haben können, sofern sie von den Konsumenten nicht als solche identifiziert, sondern für CSR-Initiativen gehalten werden (vgl. De Jong et al., 2018, S. 99). Daraus ergibt sich die Hypothese:

> *H_8: Greenwashing-Methoden haben positive Effekte auf die Unternehmensreputation, wenn sie fälschlicherweise für CSR-Initiativen gehalten werden.*

Außerdem wird erneut angenommen, dass eine visuelle Kommunikation, wie bspw. mithilfe eines Umweltlabels mehr Aufmerksamkeit erregen kann und überzeugender ist als rein schriftliche Aussagen (vgl. McQuarrie & Mick, 2003, S. 218; A. A. Mitchell, 1986, S. 16-21; Burns et al., 1993, S. 71). Der Effekt einer visuellen Greenwashing-Methode mit einem Fake-Label ist also vermutlich stärker als der Effekt einer Methode mit vagen Formulierungen. Ob der Effekt stärker positiv oder negativ ist hängt davon ab, ob die Werbeanzeige als Greenwashing oder CSR identifiziert wird. Als nächste Hypothese erschließt sich daraus:

> *H_9: Greenwashing-Methoden mit einem Fake-Label haben stärkere Effekte auf die Unternehmensreputation als solche mit vagen Formulierungen.*

3.3. Die Effekte unterschiedlicher Greenwashing-Skandale auf die Unternehmensreputation

In dieser Studie sollen nicht nur die Effekte unterschiedlicher CSR-Initiativen und Greenwashing-Methoden auf die Reputation untersucht werden, sondern auch die Effekte unterschiedlicher Greenwashing-Skandale. Das heißt, es werden auch Situationen betrachtet, in denen die irreführenden Werbeaussagen bezüglich der Umweltpraktiken eines Unternehmens aufgedeckt worden sind und darüber berichtet wird. Auch hier wird aufgrund der Signaltheorie zunächst einmal davon ausgegangen, dass Skandale, die durch unterschiedliche Greenwashing-Methoden ausgelöst wurden, unterschiedliche Reaktionen auslösen (vgl. Connelly et al., 2011, S. 39). Aufgrund dessen haben sie also unterschiedliche Effekte auf die Unternehmensreputation zur Folge, woraus sich die Hypothese ergibt:

> *H_{10}: Unterschiedliche Greenwashing-Skandale haben unterschiedliche Effekte auf die Unternehmensreputation.*

Darüber hinaus ist in der Literatur bisher erst sehr wenig über die Effekte bzw. Wirkungen von Greenwashing-Skandalen bekannt. Eines der wenigen relevanten Beispiele ist das des Abgasskandals von Volkswagen. Nachdem bekannt wurde, dass der deutsche Automobilkonzern viele Fahrzeuge manipulierte, um die gesetzlich vorgeschriebenen Grenzwerte für Autoabgase zu umgehen, folgte unter anderem ein enormer Rückgang des Aktienkurses. Dies könnte selbstverständlich auch eine Folge der beschädigten Reputation von VW sein, was jedoch nicht belegt ist. Eine Vielzahl von Faktoren, wie dem möglichen Verlust von Marktanteilen in naher Zukunft oder die finanziellen Belastungen durch Entschädigungen und Strafen könnten dabei eine weitaus größere Rolle gespielt haben (vgl. Majláth, 2016, S. 117-118).

In Abschnitt 3.2 wurden allerdings bereits die möglichen Effekte von Greenwashing identifiziert, sofern die Konsumenten erkennen, dass es sich bei den vermeintlich umweltfreundlichen Initiativen eines Unternehmens tatsächlich um Greenwashing handelt. Erhalten Konsumenten Signale, wie bspw. Informationen über einen Greenwashing-Skandal bei einem Unternehmen, aus anderen Quellen, wie den formellen Medien (z. B. Zeitungen, Fernsehen; Van den Bogaerd & Aerts, 2014, S. 28; Mason, 2014, S. 87-88) oder sozialen Medien (z. B. Blogs; Fan, Geddes & Flory, 2013, S. 114), ist davon auszugehen, dass sich diese negativen Effekte von Greenwashing noch verstärken. Daraus ergibt sich die Hypothese:

> *H_{11}: Greenwashing-Skandale verstärken die negativen Effekte von Greenwashing auf die Unternehmensreputation.*

Auch in diesem Fall erregen visuelle Kommunikationsmethoden vermutlich mehr Aufmerksamkeit als rein schriftliche Methoden (vgl. McQuarrie & Mick, 2003, S. 218; A. A. Mitchell, 1986, S. 16-21; Burns et al., 1993, S. 71). Außerdem sind vage Behauptungen zwar irreführend, aber womöglich werden sie nach einem Skandal von Konsumenten eher verziehen als gefälschte Umweltzeichen. Diese Überlegungen führen zu der Hypothese:

> *H_{12}: Greenwashing-Skandale zu einem Fake-Label haben stärkere negative Effekte auf die Unternehmensreputation als solche zu vagen Formulierungen.*

3.4. Die Einflüsse von Konsumenteneigenschaften auf die Reputation

Um die Reaktionen von Konsumenten auf (vermeintlich) umweltbewusste Werbebotschaften zu verstehen, muss berücksichtigt werden, dass diese von einer Vielzahl von Konsumenteneigenschaften beeinflusst werden. Zunächst einmal unterscheiden sich Konsumenten in ihren Reaktionen auf überzeugende nachhaltige Behauptungen je nach Ausmaß ihres Involvements, also nach ihrer Einbindung in umweltrelevante Themenfelder (vgl. Parguel et al., 2015, S. 19; Matthes, Wirth, Schemer & Pachoud, 2012, S. 129).

Die Forschung hat aufgezeigt, dass Konsumenten mit hohem Umweltinvolvement eher dazu neigen, Marken zu verurteilen, die sich weniger um Umweltfreundlichkeit bemühen (vgl. Newell et al., 1998, S. 48; Do Paço & Reis, 2012, S. 152-153). Studien haben auch ergeben, dass Konsumenten mit höherem Involvement eher irreführende Absichten in der Werbung erkennen. Dies ist ein Mechanismus, der auf ihr größeres Engagement bei der detaillierten Verarbeitung von Werbeaussagen zurückgeführt wird (vgl. Johar, 1995, S. 276-277). Bei bestimmten Werbebotschaften kann ein größeres Involvement auf der anderen Seite jedoch auch die affektive Überzeugung erhöhen (vgl. Buck, Anderson, Chaudhuri & Ray, 2004, S. 649). Ergebnisse aus verschiedenen Studien deuten darauf hin, dass ökologisch engagierte Verbraucher positiver auf grüne Werbung reagieren (vgl. Schmuck, Matthes & Naderer, 2018, S. 414; Hartmann, Apaolaza & Eisend, 2016, S. 436-437). Dies liegt vermutlich an ihrer größeren emotionalen Affinität zu Nachhaltigkeit und Umweltfreundlichkeit (vgl. Hartmann, Apaolaza & Alija, 2013, S. 203-204).

Im Green Marketing gibt es zwei Schlüsselmerkmale des Umweltinvolvements, welche die Reaktion der Konsumenten auf verschiedene Werbeaussagen beeinflussen: Umweltbewusstsein (Environmental Concern; EC) und Umweltwissen (Environmental Knowledge; EK).

1. Umweltbewusstsein (EC).
   Umweltbewusste Konsumenten werden von verschiedenen Studien als solche Konsumenten charakterisiert, die sich in hohem Maße um die Umwelt sorgen, ein außergewöhnliches Bewusstsein für Umweltprobleme haben und die Notwendigkeit des Umweltschutzes erkennen (vgl. Hartmann & Apaolaza-Ibáñez, 2012, S. 1254-1255; Matthes & Wonneberger, 2014, S. 1886-1887; Schuhwerk/Lefkoff-Hagius 1995, S. 46; Tucker, Rifon, Lee & Reece, 2012, S. 9). Weiterhin wird angenommen, dass Umweltaussagen von Unternehmen

Konsumenten mit einem höheren EC-Grad stärker beeinflussen (vgl. Grimmer & Bingham, 2013, S. 1945).

2. Umweltwissen (EK).
   Ein weiterer wichtiger Indikator für das Umweltinvolvement der Konsumenten ist ihr objektives Wissen über ökologische Themen (vgl. Ellen, 1994, S. 43; Parguel et al., 2015, S. 113-114). Die Forschung aus verschiedenen Bereichen des Marketings hat allgemein festgestellt, dass themenbezogenes Wissen ein entscheidender Moderator von Werbewirkungen und ein wichtiger Faktor bei der Beurteilung der Glaubwürdigkeit von Werbeaussagen ist; insbesondere, wenn solche Aussagen mehrdeutig sind (vgl. Andrews, Burton & Netemeyer, 2000, S. 29). Während EC als eher subjektiver Indikator das Bewusstsein der Konsumenten für Umweltbelange angibt, fungiert EK als objektiver Indikator für ihre Fähigkeit, Greenwashing wahrzunehmen (vgl. Schmuck, Matthes & Naderer, 2018, S. 131).

Basierend auf diesen Überlegungen werden die Interaktionseffekte beider Dimensionen des Umweltinvolvements, EC und EK mit CSR-Initiativen sowie mit unterschiedlichen Greenwashing-Methoden untersucht. Dabei wird angenommen, dass Umweltinvolvement im Sinne von EC und EK rationale und affektive Mechanismen verstärkt. Anders ausgedrückt wird erwartet, dass ein hohes Umweltinvolvement dazu führt, dass sowohl CSR-Initiativen als auch Greenwashing-Methoden eher als solche erkannt werden. Es ergibt sich also folgende Hypothese:

> *H_{13}: Ein hohes Umweltinvolvement führt dazu, dass CSR-Initiativen und Greenwashing-Methoden eher als solche erkannt werden.*

Zusätzlich dazu kann auch das Alter der Konsumenten deren Reaktion auf Greenwashing-Methoden beeinflussen. So wurde Skepsis gegenüber Green Marketing insbesondere unter jungen Menschen festgestellt (vgl. Aji & Sutikno, 2015, S. 460-461; Musgrove, Choi & Chris Cox, 2018, S. 285-286). Andere Studien haben bestätigt, dass junge Menschen, insbesondere Universitätsstudenten, sehr informiert über Fragen im Zusammenhang mit Umwelt und Umweltkommunikation sind und überdurchschnittlich sensibel auf diese reagieren (vgl. Besel, Burke & Christos, 2017, S. 67-70; Chan & Leung, 2006, S. 436). Für diese Studie lässt sich dadurch folgende Hypothese schlussfolgern:

> *H_{14}: Ein junges Alter führt dazu, dass CSR-Initiativen und Greenwashing-Methoden eher als solche erkannt werden.*

Bei einer Studie von Torelli et al. (2020, S. 416-417) über die Wahrnehmung von Greenwashing wurde abschließend beobachtet, dass auch das Geschlecht der Konsumenten ein wichtiger Einflussfaktor sein kann. So haben die weiblichen Probanden Greenwashing signifikant häufiger erkannt als die männlichen, woraus folgt:

> *H_{15}: Frauen können CSR-Initiativen und Greenwashing-Methoden häufiger als solche identifizieren als Männer.*

Auch bei den Reaktionen auf einen Greenwashing-/ bzw. Umweltskandal wurden Unterschiede festgestellt. Demnach schädigt ein solcher die Meinung über das Unternehmen bei Frauen langfristiger, wodurch sie das Unternehmen in Zukunft eher meiden als Männer (vgl. Torelli et al., 2020, S. 416-417). Die letzte Hypothese dieser Studie lautet demnach:

> *H_{16}: Greenwashing-Skandale haben bei Frauen stärkere negative Effekte als bei Männern.*

Tabelle 1 gibt einen Gesamtüberblick über die Hypothesen, die in dieser Studie überprüft werden.

3.5. Die Messung der Unternehmensreputation mithilfe des RepTrak-Konzepts

Die Reputation ist einer der wichtigsten Faktoren für die Wettbewerbsfähigkeit von Unternehmen (vgl. Argenti & Druckenmiller, 2004, S. 368). Aufgrund der Tatsache, dass die Grenzen zwischen Organisationen und ihren Stakeholdern in der heutigen Geschäftswelt immer weiter verschwimmen, besteht die Notwendigkeit, die Reputation von Unternehmen strategisch zu managen. Die Entwicklung und Aufrechterhaltung eines erkennbaren Images und einer guten Reputation können drüber entscheiden, ob ein Unternehmen langfristig erfolgreich sein wird oder nicht (vgl. Gray & Balmer, 1998, S. 695).

Mit dem RepTrak-Konzept wurde 2005-2006 ein Analyseinstrument entwickelt, anhand dessen die Wahrnehmung eines Unternehmens durch Stakeholder beobachtet und bewertet werden kann. Außerdem ermöglicht es ein umfassenderes Verständnis der zugrundeliegenden Treiber der Unternehmensreputation. Von besonderem Wert ist die nachgewiesene Validität des RepTrak-Konzepts, zur branchen-, stakeholder- und länderübergreifenden Messung der Reputation (vgl. C. J. Fombrun et al., 2015, S. 4). Es basiert auf der Überlegung, dass die Gesamtreputation eines Unternehmens in den Wahrnehmungen seiner Stakeholder verwurzelt ist (vgl. Newburry, 2010, S. 388), von denen jeder auf unterschiedliche Signale bzw. Informationsinputs reagiert (vgl. Spence, 1978, S. 358-359; Prabhu & Stewart, 2001, S. 70). Durch die Untersuchung der Arten von Informationsinputs, welche die Wahrnehmung eines Unternehmens durch die Stakeholder beeinflussen, können verschiedene Dimensionen erkannt werden, welche wahrscheinlich die emotionalen Reaktionen der Stakeholder in Form von Markenwahrnehmung und Markenvertrauen auslösen (vgl. C. J. Fombrun et al., 2015, S. 4).

Das RepTrak-Konzept ist auch die Grundlage des Global RepTrak 100, einer jährlichen Studie zur Identifizierung der Unternehmen mit der höchsten Reputation, welche in wichtigen Wirtschaftsmedien, wie bspw. Forbes, veröffentlicht wird (vgl. Wiedmann, 2012, S. 65-66). Das Konzept besteht aus

Tabelle 1: Überblick über die Hypothesen

Quelle: Eigene Darstellung

H_1	Unterschiedliche CSR-Initiativen haben unterschiedliche Effekte auf die Unternehmensreputation.
H_2	CSR-Initiativen haben insgesamt positive Effekte auf die Unternehmensreputation.
H_3	CSR-Initiativen haben negative Effekte auf die Unternehmensreputation, wenn sie fälschlicherweise für Greenwashing gehalten werden.
H_4	Als intrinsisch motiviert wahrgenommene CSR-Initiativen haben stärkere positive Effekte auf die Unternehmensreputation als solche, die als extrinsisch motiviert wahrgenommen werden.
H_5	CSR-Initiativen mit einem Umweltlabel haben stärkere positive Effekte auf die Unternehmensreputation als CSR-Initiativen ohne Umweltlabel.
H_6	Unterschiedliche Greenwashing-Methoden haben unterschiedliche Effekte auf die Unternehmensreputation.
H_7	Greenwashing-Methoden haben insgesamt negative Effekte auf die Unternehmensreputation.
H_8	Greenwashing-Methoden haben positive Effekte auf die Unternehmensreputation, wenn sie fälschlicherweise für CSR-Initiativen gehalten werden.
H_9	Greenwashing-Methoden mit einem Fake-Label haben stärkere Effekte auf die Unternehmensreputation als solche mit vagen Formulierungen.
H_{10}	Unterschiedliche Greenwashing-Skandale haben unterschiedliche Effekte auf die Unternehmensreputation.
H_{11}	Greenwashing-Skandale verstärken die negativen Effekte von Greenwashing auf die Unternehmensreputation.
H_{12}	Greenwashing-Skandale zu einem Fake-Label haben stärkere negative Effekte auf die Unternehmensreputation als solche zu vagen Formulierungen.
H_{13}	Ein hohes Umweltinvolvement führt dazu, dass CSR-Initiativen und Greenwashing-Methoden eher als solche erkannt werden.
H_{14}	Ein junges Alter führt dazu, dass CSR-Initiativen und Greenwashing-Methoden eher als solche erkannt werden.
H_{15}	Frauen können CSR-Initiativen und Greenwashing-Methoden häufiger als solche identifizieren als Männer.
H_{16}	Greenwashing-Skandale haben bei Frauen stärkere negative Effekte als bei Männern.

sieben Dimensionen mit insgesamt 23 Items. Diese sieben Dimensionen des RepTrak werden im Folgenden genauer erläutert.

1. Produkte

Die Produkte und Dienstleistungen sind in der Regel der größte Berührungspunkt zwischen Unternehmen und ihren Stakeholdern. Insbesondere die Konsumenten kennen Unternehmen vor allem durch die Produkt- und Dienstleistungsangebote auf dem Markt. Dementsprechend kann davon ausgegangen werden, dass die Reputation eines Unternehmens durch die Wahrnehmung seiner Produkte beeinflusst wird (vgl. Veloutsou & Moutinho, 2009, S. 315; Rao, Qu & Ruekert, 1999, S. 266; K. T. Smith, Smith & Wang, 2010, S. 12). Dazu zählen unter anderem die Qualität, der Preis, zu dem es verkauft wird, der wahrgenommene Wert, der angebotene Kundensupport und der Glaube an die Bereitschaft des Unternehmens, hinter seinen Produkten und Dienstleistungen zu stehen (vgl. Dawar & Parker, 1994, S. 91).

Auch Investitionen, bspw. in neue Herstellungsmethoden oder Materialien für die Produkte eines Unternehmens können Veränderungen erzeugen, inwieweit die Stakeholder ein Unternehmen als bewundernswert, sympathisch oder vertrauenswürdig wahrnehmen. Spieltheoretische Modelle gehen davon aus, dass die Reputation eines Unternehmens vor allem durch Investitionen zur Steigerung der Produktqualität aufgebaut wird (vgl. Milgrom & Roberts, 1986, S. 819-820).

Zusätzlich dazu haben Pritchard und Wilson (2018, S. 47) in Übereinstimmung mit früheren Arbeiten (vgl. Olsen, Slotegraaf & Chandukala, 2014, S. 119) herausgefunden, dass die Einführung neuer umweltfreundlicher Produkte die Wahrnehmung eines Unternehmens verändern kann. Die positiven Attribute, die Konsumenten in diesen Produkten erkennen, tragen demnach dazu bei, dass sich die Reputation von Unternehmen verbessert. Die RepTrak-Dimension „Produkte/Dienstleistungen“ konzentriert sich also auf Aspekte wie Produktqualität, das Preis-Leistungs-Verhältnis sowie die Bedürfnisse und Erwartungen der Stakeholder (vgl. C. J. Fombrun et al., 2015, S. 6).

2. Innovation

Für viele Firmen ist die Wahrnehmung als innovatives Unternehmen etwas, das sowohl geschätzt als auch aktiv angestrebt wird. Als wichtiger Unternehmenswert (vgl. Fang, Palmatier & Grewal, 2011, S. 587) bezieht sich Innovation auf einen Verbesserungs- und Modernisierungsaspekt und erzeugt daher eine emotionale Reaktion der Bewunderung für den Innovator (vgl. C. J. Fombrun et al., 2015, S. 6). Die empirischen Ergebnisse deuten darauf hin, dass ein hoher von den Konsumenten wahrgenommener Innovationsfaktor zu Begeisterung und erhöhter Loyalität gegenüber dem innovativen Unternehmen führt. Als weitere Ergebnisse wurden ein positiveres Gesamtimage des Unternehmens und eine höhere Toleranz gegenüber gelegentlichen Produktfehlern festgestellt (vgl. Henard & Dacin, 2010, S. 330). Auch Höflinger, Nagel und Sandner (2018, S. 36) zeigen, dass die Innovationsleistung von Unternehmen signifikant mit der Reputation für technologische Innovationen verknüpft ist. So zeigt ihre Studie, dass die erfinderische Leistung mit der Unternehmensreputation korreliert. Zusammengefasst bestätigt die Forschung eine Beziehung zwischen Innovation und Reputation, welche oft von der effektiven Kommunikation über eine Innovation abhängt (vgl. Courtright & Smudde, 2009, S. 262). Unternehmen, die sich schnell an Veränderungen anpassen, neue Produkte auf den Markt bringen und neue Ideen entwickeln, ernten eher Respekt und Bewunderung. Erkennbar wird die Innovationskraft von Unternehmen durch Rankings von Fachzeitschriften wie Bloomberg oder Forbes. Diese Publikationen signalisieren allen Beobachtern die Innovationskraft eines Unternehmens und tragen damit zu dessen Reputation bei. Die Dimension „Innovation“ des RepTrak-Konzepts bewertet die Wahrnehmung eines Unternehmens als innovativ und anpassungsfähig (vgl. C. J. Fombrun et al., 2015, S. 6).

3. Arbeitsumfeld

Die meisten Stakeholder mögen und respektieren solche Unternehmen, die ein kollegiales Umfeld bieten und für Chancengleichheit unter den Mitarbeitern sorgen. Forschungsergebnisse zeigen, dass die Reputation und die finanzielle Performance unter anderem durch die Arbeitszufriedenheit beeinflusst werden (vgl. Carmeli & Freund, 2002, S. 51). Dem entgegengesetzt kann die Unzufriedenheit der Mitarbeiter die Chancen eines Unternehmens verschlechtern, positive Beziehungen sowie einen guten Ruf in der Gesellschaft aufzubauen und zu erhalten (vgl. Duffy, Ganster & Pagon, 2002, S. 333).

Die Literatur geht davon aus, dass sich zufriedene Mitarbeiter langfristiger und intensiver im Unternehmen engagieren, das Unternehmen seltener wechseln und daher häufiger als Botschafter des Unternehmens auftreten. Außerdem ist die Wahrnehmung eines Unternehmens als guter Arbeitsplatz entscheidend für die Rekrutierung hochwertiger Arbeitskräfte (vgl. Alniacik, Alniacik & Erdogmus, 2012, S. 3; Nolan, Gohlke, Gilmore & Rosiello, 2013, S. 308-309). Verschiedene Fachzeitschriften beurteilen Unternehmen danach, wie sie ihre Mitarbeiter behandeln. Bspw. veröffentlicht Fortune regelmäßig „The 100 Best Companies to Work For“ und Forbes eine Liste mit „The 25 Best Places to Work“. Beide machen damit Informationen über das Arbeitsumfeld sichtbar (vgl. C. J. Fombrun et al., 2015, S. 6-7). Solche Informationen sagen aus, wie fair ein Unternehmen seine Mitarbeiter behandelt, erzeugen dadurch Vertrauen und Respekt bei den Stakeholdern und tragen so zum Aufbau einer positiven Reputation für das Unternehmen bei. Zusammengefasst bewertet die RepTrak-Dimension „Arbeitsplatz“ die Wahrnehmung der Methoden eines Unternehmens, ein Umfeld zu schaffen, welches sich um die Mitarbeiter kümmert und sie fair und gerecht behandelt (vgl. C. J. Fombrun et al., 2015, S. 6-7).

4. Leitung

Die Leitung, oder auch Governance, steht für den Grad der Offenheit und Transparenz eines Unternehmens und bezieht sich auch auf das ethische Verhalten und die Art und Weise, wie das Unternehmen seine Geschäfte führt (vgl. Wiedmann, Hennigs, Schmidt & Wuestefeld, 2013, S. 192). Es kann im weitesten Sinne als die Analyse von Macht und Einfluss in Bezug auf die Entscheidungsfindung innerhalb eines Unternehmens definiert werden (vgl. Aguilera & Jackson, 2010, S. 487).

G. F. Davis (2005, S. 143) erklärt Corporate Governance als die Strukturen, Prozesse und Institutionen innerhalb und im Umfeld von Organisationen, die Macht und Ressourcenkontrolle unter den Beteiligten verteilen. In Anbetracht der zunehmenden Komplexität multinationaler Unternehmen wird die Leitung ebendieser zunehmend als Schlüsselthema anerkannt (vgl. Ghosh & John, 2009, S. 609; S. K. Kim, McFarland, Kwon, Son & Griffith, 2011, S. 603) und professionelle Leitungsstrukturen gelten als eine wichtige Komponente des Reputationsmanagements (vgl. Casado, Peláez & Cardona, 2014, S. 46). So wurde unter anderem festgestellt, dass eine transparente Unternehmensleitung positive Effekte auf die Wirksamkeit der CSR-Initiativen eines Unternehmens hat (vgl. Jo & Harjoto, 2012, S. 53). Stakeholder sind regelmäßig Informationen über die Leitung eines Unternehmens ausgesetzt, bspw. durch die Medien, Wirtschaftsprüfer oder Regierungsbehörden. Je mehr ein Unternehmen als ethisch und transparent wahrgenommen wird, desto wahrscheinlicher ist es, dass es in den Köpfen der meisten Stakeholder Bewunderung und Vertrauen erzeugt - und damit positive Reputation (vgl. Soleimani, Schneper & Newburry, 2014, S. 991). Unternehmen selbst unterzeichnen oft institutionelle Verhaltenskodizes (bspw. den Deutschen Corporate Governance Kodex), um den Stakeholdern ihre Prinzipien und Verpflichtungen zu signalisieren und Vertrauen dafür zu schaffen, dass ihre internen Praktiken verlässlich sind. Die Dimension „Leitung“ von RepTrak bewertet somit die Wahrnehmung eines Unternehmens durch Stakeholder als ethisch, fair und transparent (vgl. C. J. Fombrun et al., 2015, S. 7).

5. Gesellschaftliche Einbindung

Unternehmen sollten nicht nur profitable Werkzeuge sein, sondern sich auch ihrer sozialen Verantwortung bewusst sein. Unternehmenspraktiken benötigen umfassende Strukturen, um über auferlegte Standards hinauszugehen und Ziele für eine nachhaltige Entwicklung zu erreichen (vgl. Tai, Chuang et al., 2014, S. 117). Die Perspektive und Umset-

zung von CSR hat wichtige Konsequenzen für multinationale Unternehmen. Qualitative Studien legen nahe, dass Stakeholder dazu neigen, Unternehmen für ihren gesellschaftlichen Mehrwert zu respektieren und zu bewundern (vgl. Mishra & Suar, 2010, S. 585; Orlitzky & Swanson, 2012, S. 133-134). Darüber hinaus deuten empirische Belege darauf hin, dass die gesellschaftliche Einbindung ein legitimitätsbildender strategischer Vermögenswert ist (vgl. Sridhar, 2012, S. 74-75), der zu verschiedenen Formen der Loyalität gegenüber einem Unternehmen führen (vgl. Aaron, McMillan & Cline, 2012, S. 304) und sogar einen Puffer darstellen kann, der Unternehmen in Krisenzeiten schützt (vgl. Mio & Fasan, 2012, S. 281). Dies hat zur Folge, dass viele Unternehmen erhebliche Geldbeträge und Marketinganstrengungen aufwenden, um ihre gesellschaftliche Einbindung zu erhöhen und als nachhaltiges, sozialverantwortliches Unternehmen wahrgenommen zu werden (vgl. Gottschalk, 2013, S. 178-180; Morris, Bartkus, Glassman & Rhiel, 2013, S. 285-286; Vlachos, Krepapa, Panagopoulos & Tsamakos, 2013, S. 248). Empirisch ist eine nachhaltige Unternehmensleistung eines der wichtigsten Korrelationselemente der Unternehmensreputation (vgl. Lange, Lee & Dai, 2011, S. 157).). Allerdings hat sich auch dadurch, dass immer mehr Fälle von Greenwashing bei Unternehmen öffentlich werden, eine allgemeine Skepsis und Abneigung gegenüber den Behauptungen von Unternehmen zur sozialen Verantwortung entwickelt (vgl. Connors, Anderson-MacDonald & Thomson, 2017, S. 599; Koschate-Fischer, Stefan & Hoyer, 2012, S. 921). Die Dimension „gesellschaftliche Einbindung“ bezieht sich auf den positiven Einfluss auf die Gesellschaft, den ein Unternehmen den Stakeholdern bietet und untersucht die Wahrnehmung eines Unternehmens als umweltfreundlich, gemeinnützig und als positiver Einfluss auf die Gesellschaft (vgl. C. J. Fombrun et al., 2015, S. 7-8).

6. Führung

Auch das Management und die Führung eines Unternehmens haben einen Einfluss auf seine wahrgenommene Reputation. Für viele Wissenschaftler ist visionäre Führung essenziell für die Wettbewerbsfähigkeit eines Unternehmens (vgl. Kantabutra & Avery, 2010, S. 37). CEOs können wichtige Impulsgeber sein, um Bewunderung und Vertrauen bei Stakeholdern zu erzeugen (vgl. Flatt, Harris-Boundy & Wagner, 2013, S. 213; Halff, 2013, S. 240-141). Studien über Celebrity-CEOs (vgl. Treadway, Adams, Ranft & Ferris, 2009, S. 565), Star-CEOs (vgl. Wade, Porac, Pollock & Graffin, 2008, S. 207) und CEO Brands (vgl. Bendisch, Larsen & Trueman, 2013, S. 609-610) bekräftigen die Bedeutung von Führungskräften bei der Vermittlung des Erfolgs und der Leistung eines Unternehmens gegenüber den Konsumenten und anderen Stakeholdern. Moderne Beispiele dafür sind Bill Gates, Larry Page oder Elon Musk. Die Beliebtheit dieser Personen basiert auf ihrem Erfolg und der Tatsache, dass ihnen eine Reihe ausgeprägter Talente nachgesagt werden, welche die Erwartungen von Stakeholdern auf einzigartige Weise befriedigen (vgl. Ulrich & Smallwood, 2007, S. 95; Cottan-Nir, 2019, S. 121).

Zudem bestätigt die Forschung, dass sich Führungskräfte in ihren strategischen Fähigkeiten unterscheiden (vgl. Goldfarb & Yang, 2009, S. 621). Visionäre Führungspersönlichkeiten ziehen eine positive Medienberichterstattung und die Unterstützung von Investoren an und signalisieren so allen Stakeholdern die Glaubwürdigkeit der Unternehmensaktivitäten. Darüber hinaus erhöhen sie das Vertrauen in das Unternehmen und bauen so eine positive Unternehmensreputation auf (vgl. Westphal & Deephouse, 2011, S. 1073). Die RepTrak-Dimension „Führung“ untersucht die Wahrnehmung von Führungskräften und Visionen als wichtige Antreiber eines Unternehmens (vgl. C. J. Fombrun et al., 2015, S. 8).

7. Performance

Die finanzielle Performance gibt Aufschluss über den wirtschaftlichen Erfolg eines Unternehmens, also sowohl die aktuelle Profitabilität als auch das Potential für zukünftiges Wachstum. Insbesondere für Investoren sind diese Informationen über den operativen Erfolg wichtige Entscheidungshilfen. Die positiven Effekte einer guten Unternehmensreputation auf die finanzielle Performance sind in der Literatur umfassend beschrieben (vgl. Roberts & Dowling, 2002, S. 1077; J. P. Walsh, Weber & Margolis, 2003, S. 867; Eberl & Schwaiger, 2005, S. 851). Darüber hinaus hat sich gezeigt, dass Rentabilität und Wachstumsaussichten die Bewertungen großer Unternehmen beeinflussen und stark mit der Reputation korrelieren (vgl. C. Fombrun & Shanley, 1990, S. 233). Performance steht also für Erfolg und starke Wachstumsaussichten. Die Dimension „Performance“ des RepTrak-Konzepts basiert daher auf Attributen, welche die Wahrnehmung der Stakeholder in Bezug auf die finanzielle Gesamtleistung, die Profitabilität und die Wachstumsaussichten eines Unternehmens bewerten (vgl. C. J. Fombrun et al., 2015, S. 8).

Anhand dieser sieben Dimensionen mit insgesamt 23 Items sollen in dieser Studie die Effekte unterschiedlicher CSR-Initiativen, Greenwashing-Methoden und Greenwashing-Skandale auf die Unternehmensreputation am Beispiel der Marke Adidas untersucht werden.

3.6. Grafische Darstellung und Erläuterung des konzeptionellen Bezugsrahmens

Im Folgenden sollen die theoretischen Ansätze für die Effekte unterschiedlicher CSR-Initiativen, Greenwashing-Methoden sowie Greenwashing-Skandale auf die Unternehmensreputation grafisch dargestellt werden.

Um mögliche Einflüsse aufgrund unterschiedlicher Voreinstellungen auszuschließen, muss vor dem Kommunikationsmedium die Einstellung gegenüber der zu betrachtenden Marke überprüft werden. Im Zentrum dieser Untersuchung steht der Stimulus, also die unterschiedlichen Werbeanzeigen, welche womöglich unterschiedliche Effekte auf die Reputation der Marke Adidas haben. Als CSR-Initiativen werden zum einen ein Produkt mit einem zertifiziertem Umweltlabel und zum anderen mit einem umfangreichen Hinweistext zur nachhaltigen Herstellung des Produkts gezeigt. Im Gegensatz dazu ist bei der ersten Greenwashing-Methode ein gefälschtes, also nicht zertifiziertes Umweltlogo zu sehen und bei der zweiten werden vage, irreführende Aussagen

verwendet. Die CSR-Initiativen bilden also gewissermaßen positive Gegenbeispiele in Bezug auf die beiden verwendeten Greenwashing-Methoden. Die Greenwashing-Anzeigen werden auch für die Greenwashing-Skandale verwendet. Darüber hinaus werden hier jeweils passende Zeitungsartikel gezeigt, in denen die Greenwashing-Methode dargestellt und erläutert wird. Die genaue Darstellung der Stimuli sowie der gesamte Aufbau und die verwendeten Skalen des Experiments werden in Kapitel 4.1 noch genauer erläutert. Zur Messung der Reputation wird das RepTrak-Konzept verwendet. Dieses wurde aus Studien des Reputation Institute entwickelt, um ein systematisches Instrument zur Erfassung und Analyse von Stakeholder-Wahrnehmungen bereitzustellen, welches Unternehmen helfen kann, ihre Reputation und ihre Auswirkungen auf das Stakeholder-Verhalten besser zu kontrollieren (vgl. C. J. Fombrun et al., 2015, S. 3). Alle 23 Items der sieben Dimensionen werden dabei berücksichtigt.

Die Reputation ist die Wahrnehmung einer Marke aus Perspektive der Stakeholder, in dieser Studie insbesondere der Konsumenten. Um diese Wahrnehmung noch genauer messen zu können, wird der Brand Potential Index (BPI) als Instrument zur Markenbewertung verwendet. Der BPI misst branchenübergreifend die Attraktivität einer Marke als psychische Markenstärke aus der Kundenperspektive und ist dabei mit dem Einkaufverhalten verknüpft. Er lässt sich in insgesamt drei Kategorien und zehn Items unterteilen (vgl. Högl & Hupp, 2004, S. 130; Grimm, Högl & Hupp, 2000, S. 4-18; Esch, Langner & Brunner, 2005, S. 1242). Da die Kategorie „rationale Wertschätzung“ bereits umfangreich durch das RepTrak-Konzept abgedeckt ist, wird sich hier auf die Markenwahrnehmung (MW) sowie das Markenverhalten (MV) beschränkt. Die Markenwahrnehmung umfasst die Items „Markenimage“, „Markenvertrauen“ sowie „Markenidentifikation“.

Nun ist aus einer Vielzahl wissenschaftlicher Studien bekannt, dass sich die Wahrnehmung einer Marke auf das Verhalten der Konsumenten auswirkt. Ein von den Kunden wahrgenommenes hohes Image lässt bspw. oftmals eine höhere Kaufbereitschaft folgen (vgl. Faircloth, Capella & Alford, 2001, S. 61-62; Vázquez, Del Rio & Iglesias, 2002, S.40-43). Weiterhin wirkt sich die Markenbekanntheit intensiv auf den Entscheidungsprozess der Konsumenten aus (vgl. Huang & Sarigöllü, 2014, S. 112; Macdonald & Sharp, 2000, S. 12-13) und die wahrgenommene Qualität beeinflusst maßgeblich die Einstellungen und Kaufabsichten der Konsumenten (vgl. Pappu, Quester & Cooksey, 2005, S. 151). Auch J. Kim und Lennon (2013, S. 33) haben herausgefunden, dass die Reputation einer Marke einen signifikant positiven Effekt auf die Emotionen der Konsumenten und einen signifikant negativen Effekt auf das wahrgenommene Risiko hat. Dadurch konnte ein deutlicher Anstieg der Kaufabsicht beobachtet werden. Zusammengefasst besteht also ein enger Zusammenhang zwischen den psychischen und den erlösorientierten Indikatoren. Marken mit einer hohen Markenwahrnehmung weisen demnach häufig auch einen hohen ökonomischen Erfolg auf. Aus diesem Grund wird auch die dritte Kategorie des BPI betrachtet, das Markenverhalten. Diese beinhaltet die Items „Kaufbereitschaft“, „Weiterempfehlungsabsicht“ sowie die „Mehrpreisakzeptanz“.

Bei der Untersuchung der Effekte der unterschiedlichen Werbeanzeigen wird darüber hinaus die wahrgenommene Motivation der Marke und die Wahrnehmung von Greenwashing analysiert. Es wird angenommen, dass die CSR-Initiativen unterschiedlich starke Effekte haben, je nachdem, ob die Konsumenten die Initiative eher als strategisch oder als ökologisch motiviert wahrnehmen (vgl. Graafland & Van de Ven, 2006, S. 121). Weiterhin ist die Wahrnehmung von Greenwashing entscheidend. Aufgrund der zunehmenden Konsumentenskepsis gegenüber Green Marketing kann es vorkommen, dass tatsächliche CSR-Initiativen als Greenwashing-Methoden wahrgenommen werden (vgl. Jahdi & Acikdilli, 2009, S. 111). In dem Fall sind deutlich schlechtere Effekte auf die Markenreputation zu erwarten. Auf der anderen Seite kann auch eine Greenwashing-Methode fälschlicherweise als CSR-Initiative wahrgenommen werden, wodurch die Effekte wahrscheinlich deutlich positiver ausfallen.

Bei der Betrachtung der Greenwashing-Skandale bedarf es keiner Untersuchungen bezüglich der Wahrnehmung von Greenwashing sowie der wahrgenommenen Motivation der Marke, da die irreführenden Handlungen von Adidas hier unmissverständlich aufgedeckt werden. Stattdessen werden noch die Reaktionen auf den jeweiligen Greenwashing-Skandal betrachtet, wobei unter anderem der Vertrauensverlust sowie die Bereitschaft zur Neubewertung untersucht werden.

Essenziell sind sicherlich auch die Konsumenteneigenschaften. Insbesondere das Alter und Geschlecht sowie das Umweltinvolvement konnten als potenzielle Faktoren ausgemacht werden, welche die Effekte der unterschiedlichen Werbeanzeigen auf die Markenreputation beeinflussen können.

In dieser Studie nicht berücksichtigt werden die Einflüsse des Geschäftsfeldes. Verschiedene Studien haben herausgefunden, dass das Geschäftsfeld, in dem das betrachtete Unternehmen, also in diesem Fall Adidas, agiert, Einflüsse auf die Effekte von CSR-Initiativen und Greenwashing-Methoden haben kann (vgl. Cho et al. 2006, S. 147-148; Patten, 2002, S. 772). Unternehmen, die in umweltsensiblen Branchen (Environmentally Sensitive Industries, ESI) tätig sind, neigen demnach dazu, besonders angestrengt über ihre (vermeintliche) Umweltfreundlichkeit und Nachhaltigkeit zu berichten (vgl. Pled & Iatridis, 2012, S. 61). Umweltsensible Branchen sind bspw. Öl- und Gasförderung, Tabakproduktion, Schwerindustrie und die Papierherstellung (vgl. Jahdi & Acikdilli, 2009, S. 111). ESI-Firmen mit schlechter Umweltleistung können irreführende Umweltkommunikation betreiben, um dem Verdacht eines negativen Umwelteinflusses entgegenzuwirken und die Wahrnehmung und das Vertrauen in das Unternehmen zu verbessern. Aus diesem Grund sind Konsumenten gegenüber umweltbezogener Werbung von ESI-Unternehmen häufig noch skeptischer (vgl. Cho, Phillips, Hageman & Patten, 2009, S. 949). Zudem haben Torelli et al. (2020, S. 416) herausgefunden, dass die Reaktionen der Konsumenten si-

gnifikant stärker sind, wenn Greenwashing-Methoden bei einem Unternehmen aufgedeckt werden, welches in einer umweltsensiblen Branche tätig ist. Dies deutet darauf hin, dass Konsumenten sensibler auf Umweltschäden reagieren, wenn eine größere potenzielle Gefahr für Mensch und Umwelt besteht (bspw. durch Gaslecks, Grubenunglücke etc.).

Zusammengefasst zeigt die Abb. 2 den konzeptionellen Bezugsrahmen dieser Studie.

4. Empirische Untersuchung

4.1. Aufbau des Experiments

Um die im vorangegangenen Abschnitt aufgestellten Hypothesen zu überprüfen, wird ein Feldexperiment in Form einer Onlinebefragung konzipiert. Der gesamte Fragebogen ist im Anhang (vgl. S. A1-A23) dargestellt. Ein Experiment ist per Definition eine statistische Untersuchung mithilfe von Versuchs- und Kontrollstichproben. Ziel eines Experiments ist das Erkennen von Ursache-Wirkungs-Zusammenhängen (vgl. Raab, Unger & Unger, 2009, S. 192-195), also, ob die Veränderung einer unabhängigen (exogenen) Variable zu einer Veränderung abhängiger (endogener) Variablen führt (vgl. Kuß, Wildner & Kreis, 2014, S. 141). Untersucht wurden hierbei die Effekte unterschiedlicher CSR-Initiativen, Greenwashing-Methoden sowie Greenwashing-Skandale als unabhängige Variablen auf die wahrgenommene Reputation der Marke Adidas als abhängige Variable.

Adidas wird aus mehreren Gründen als die zu betrachtende Marke ausgewählt. Die Messung der Reputation mithilfe des RepTrak-Konzepts erfordert es, dass die Marke in dem betrachteten Teilnehmerkreis sehr bekannt ist. Die Adidas Group ist ein im DAX notierter deutscher Sportartikelhersteller und nach Nike der zweitgrößte Konzern in dieser Branche. Mit einem erwirtschafteten Umsatz von insgesamt rund 23,64 Milliarden Euro weltweit im Jahr 2019 gehört Adidas zu einer der bekanntesten Marken in Deutschland (vgl. Adidas, 2019). Aufgrund der Annahme, dass überwiegend junge Probanden an dem Experiment teilnehmen und um sowohl Frauen als auch Männer gleichermaßen anzusprechen werden weiße Sneakers, genauer gesagt weiße Adidas untersucht.

Im Folgenden werden der Aufbau der Befragung erläutert und die verwendeten Skalen aufgezeigt. Zu Beginn werden einige Einstiegsfragen zur Häufigkeit des Schuhkaufs und der favorisierten Schuhart gestellt. Darüber hinaus wird gefragt, ob die Probanden bereits Schuhe der Marke Adidas besitzen (vgl. Anhang S. A2). Anschließend, also ebenfalls noch vor dem Stimulus, wird die Einstellung des Probanden gegenüber der Marke Adidas gemessen. Mithilfe dieser Information können mögliche Unterschiede zwischen den Untersuchungsgruppen in Bezug auf die Voreinstellung erkannt werden (vgl. Anhang S. A3-A4). Zur Messung werden drei Items von Sengupta et al. (2002, S. 43) statt als semantisches Differenzial in Form einer Likert-Skala verwendet und um zwei Items von Moon (2000, S. 328) ergänzt. Tabelle 25 (vgl. Anhang S. A24-A27) liefert einen Gesamtüberblick über die verwendeten Skalen und Items.

Anschließend teilt ein Zufallstrigger die Probanden einer von sieben Gruppen zu, wobei in jeder dieser Gruppen eine andere Werbeanzeige präsentiert wird. Der Trigger strebt eine Gleichverteilung der Zuordnung an, um die Größe der Stichproben ähnlich zu halten und so die Aussagekraft zu verbessern. Die erste Gruppe ist die Kontrollgruppe. Die Probanden in dieser Gruppe bekommen eine Werbeanzeige ohne Umweltsiegel oder umweltfreundliche Informationen zu sehen. Es wird lediglich der Schuh zusammen mit dem Adidas-Logo gezeigt und erläutert, Adidas würde einen neuen Lederschuh mit verbesserter Qualität auf den Markt bringen (vgl. Anhang S. A4). Der zweiten Gruppe wird eine Werbeanzeige präsentiert, auf der neben dem Schuh und dem Logo noch ein nicht zertifiziertes Umweltlabel abgebildet ist. Darüber hinaus wird in dem Informationstext explizit auf das Nachhaltigkeitsproramm der Marke Adidas aufmerksam gemacht, welches die angebliche Auszeichnung mit dem Label „Eco Friendly“ möglich gemacht hat (vgl. Anhang S. A5). Die dritte Gruppe erhält fast genau die gleiche Werbeanzeige wie die zweite, mit dem Unterschied, dass das Umweltlabel hier zertifiziert ist. Genauer gesagt wird das Zeichen „Der blaue Engel“ neben dem Produkt dargestellt (vgl. Anhang S. A5). Die Probanden der vierten Gruppe bekommen zunächst exakt dieselbe Werbeanzeige sowie denselben Informationstext wie die zweite Gruppe gezeigt. Anschließend wird Ihnen jedoch noch ein Zeitungsartikel mit dem Titel „Greenwashing-Skandal bei Adidas“ präsentiert. In dem Artikel wird beschrieben, dass Adidas ein erfundenes Fake-Label ohne jegliche Aussagekraft verwendet hat, um für den neuen Adidas Superstar und das eigene Nachhaltigkeitsprogramm zu werben. Nun würden Experten der Marke Greenwashing vorwerfen (vgl. Anhang S. A6).

Das gleiche Muster wie bei den Gruppen zwei, drei und vier wird auch bei den Gruppen fünf, sechs und sieben verwendet, allerdings unterscheidet sich die Greenwashing-Methode und entsprechend auch die CSR-Initiative. Statt eines Fake-Labels wird den Probanden der Gruppe fünf dieselbe Werbeanzeige wie der Kontrollgruppe gezeigt, allerdings werden in dem Informationstext vage, bzw. irreführende Formulierungen verwendet. Demnach besteht der Lederschuh aus „100% natürlichen Materialien. Dies sei ein wichtiges Ziel des konzernweiten Nachhaltigkeitsprogramms gewesen.“ (Anhang S. A7). Auch den Probanden der sechsten Gruppe wird dieselbe Werbeanzeige wie der Kontrollgruppe gezeigt. Im Gegensatz zur fünften Gruppe ist im Informationstext allerdings genau beschrieben, weshalb der Schuh nachhaltig ist. Statt vager Formulierungen wird detailliert erklärt, dass der Schuh „vollständig aus Materialien, die ohne gesundheitsgefährdende Chemikalien und unter Einhaltung hoher Umweltstandards“ (Anhang S. A7) hergestellt worden ist. Abschließend wird den Probanden der Gruppe sieben zunächst dieselbe Werbeanzeige und derselbe Informationstext wie der fünften Gruppe gezeigt. Dazu bekommen auch diese Probanden noch einen Zeitungsartikel zu sehen, welcher die vagen Aussagen von Adidas kritisiert und der Marke Greenwashing vorwirft (vgl. Anhang S. A8). Zusammengefasst gibt es also sieben Gruppen: Eine Kontrollgruppe, zwei

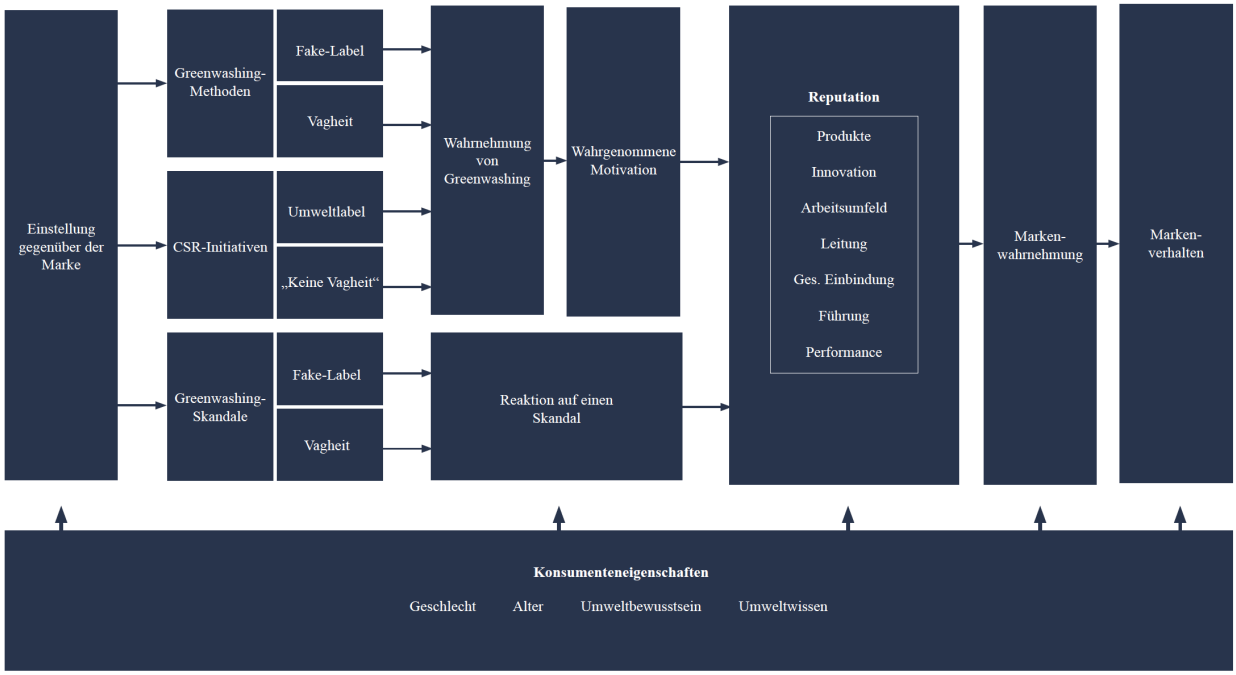


Abbildung 5: Konzeptioneller Bezugsrahmen

Quelle: Eigene Darstellung

Gruppen mit unterschiedlichen CSR-Initiativen, zwei mit unterschiedlichen Greenwashing-Methoden und zwei Gruppen mit Greenwashing-Skandalen. Die CSR-Initiativen sind dabei so entworfen worden, dass sie die jeweilige Greenwashing-Methode „ausgleichen“. Statt eines Fake-Labels wird ein zertifiziertes Umweltlabel verwendet und statt vager Formulierungen wird der Nachhaltigkeitsaspekt des Sneakers detailliert beschrieben. Die Tabelle 2 zeigt eine kurze Übersicht über die unterschiedlichen Gruppen.

Tabelle 2: Überblick über die Stichproben

Quelle: Eigene Darstellung

Gruppe Nummer	Beschreibung
1	Kontrollgruppe
2	Greenwashing Label
3	CSR Label
4	Greenwashing-Skandal Label
5	Greenwashing Vagheit
6	CSR „Keine Vagheit“
7	Greenwashing-Skandal Vagheit

Nun wird die Reputation der Marke Adidas mithilfe des RepTrak-Konzepts gemessen. Dazu werden alle 23 Items der sieben Dimensionen nach C. J. Fombrun et al. (2015, S. 15) mittels fünfstufiger Likert-Skalen (1 = stimme ganz und gar nicht zu; 5 = stimme voll und ganz zu) abgefragt (vgl. Anhang S. A9-A14). Darüber hinaus wird der BPI betrachtet. Zunächst wird die Markenwahrnehmung (Image; Vertrauen; Identifikation) und anschließend das Markenverhalten mit den erlösorientierten Indikatoren (Kaufabsicht; Weiterempfehlungsabsicht; Mehrpreisakzeptanz) abgefragt (vgl. Anhang S. A15-A16). Dafür werden die Items des BPI (vgl. Högl & Hupp, 2004, S. 130) verwendet. Im nächsten Schritt wird die wahrgenommene Motivation für das Nachhaltigkeitsprogramm der Marke Adidas untersucht. Die Probanden werden gefragt, inwieweit sie dieses als wirtschaftlich/strategisch bzw. als ökologisch/sozial motiviert wahrnehmen (vgl. Anhang S. A18). Weiterhin wird geprüft, ob die Probanden die Greenwashing-Methoden auch als solche erkennen, oder fälschlicherweise als CSR-Initiativen wahrnehmen, sowie, ob die CSR-Initiativen als solche erkannt oder aufgrund von hoher Skepsis gegenüber Green Marketing als Greenwashing-Methoden wahrgenommen werden (vgl. Anhang S. A19-A20). Dazu werden vier Items von Schmuck, Matthes und Naderer (2018, S. 134) mit zwei weiteren Items von Torelli et al. (2020, S. 412) kombiniert. Sowohl die Fragen zur Motivation der Marke als auch die zur Wahrnehmung von Greenwashing werden dabei nur den Probanden der Gruppen zwei, drei, fünf und sechs gezeigt. Die Probanden der Gruppen drei und sieben werden stattdessen zu ihrer Reaktion auf den Greenwashing-Skandal befragt (vgl. Anhang S. A17-A18). Um diese zu messen wird auf eine weitere Skala von Torelli et al. (2020, S. 412)) zurückgegriffen.

Im Anschluss daran werden die themenbezogenen Konsumenteneigenschaften, genauer gesagt das Umweltinvolvement abgefragt (vgl. Anhang S. A20-21). Für das Umweltbewusstsein werden dafür die drei Items der Skala von Schmuck, Matthes, Naderer und Beaufort (2018, S. 143)

verwendet. Das Umweltwissen wird mithilfe von vier verschiedenen Umweltlabels untersucht. Zwei dieser Label sind zertifiziert, während die anderen beiden keine Aussagekraft haben. Die Probanden werden dazu aufgefordert, die zertifizierten Label zu kennzeichnen. So kann herausgefunden werden, wie gut sich die Probanden mit häufig verwendeten Umweltlabels auskennen (vgl. Anhang S. A21). Abschließend folgen noch die Soziodemografika. Hier werden die Probanden nach ihrem Geschlecht, Alter, höchsten Bildungsabschluss, Beschäftigungsverhältnis sowie Einkommen befragt (vgl. Anhang S. A22-23).

4.2. Datenanalytische Auswertung der Untersuchungsergebnisse

4.2.1. Deskriptive Auswertung

Das Experiment dieser Arbeit ist mit „Unipark“, dem wissenschaftlichen Befragungstool von „Questback“ durchgeführt worden. Der Link zu der Umfrage wurde über diverse soziale Netzwerke verteilt. Vom 16. Januar 2021 bis 07. Februar 2021 sind insgesamt 517 Fragebögen ausgefüllt worden. Die befragten Personen sind dabei in sieben verschiedene Stichproben mit je 66-78 Probanden (vgl. Anhang Tab. 26) aufgeteilt worden, wobei jeder Stichprobe jeweils eine unterschiedliche Werbeanzeige präsentiert wurde (vgl. Anhang S. A1-A23).

Um die Stichprobe besser beschreiben zu können folgt zunächst eine kurze Erläuterung der wichtigsten Eigenschaften. Die Tabelle 3 gibt Aufschluss über die persönlichen Merkmale der Probanden. Von den 517 Personen, welche die Umfrage abgeschlossen haben, sind 262 weiblich (50,7%) und 255 männlich (49,3%), die Geschlechterverteilung ist also sehr ausgeglichen. Weiterhin handelt es sich bei den Probanden überwiegend um eher junge Personen. So beträgt das Durchschnittsalter 28 Jahre, mehr als 72% der Befragten sind unter 30 Jahre alt. Lediglich knapp 19% sind 40 Jahre oder älter. Als höchster Bildungsabschluss ist hauptsächlich das Abitur (42,9%), gefolgt vom (Fach) Hochschulabschluss (36,2%) angegeben worden. Aufgrund des niedrigen Durchschnittsalters wenig überraschend befinden sich mit 37,1% die meisten Probanden im Studium, 31,9% sind Angestellte. Das Haushaltsnettoeinkommen ist vergleichsweise hoch. Nur ein Drittel der Probanden hat weniger als 2000€ zur Verfügung. 17,6% haben hier keine Angabe gemacht.

Darüber hinaus kann die Stichprobe noch anhand dreier weiterer Indikatoren beschrieben werden. So haben die meisten Probanden (66%) angegeben, ein bis drei Paar Schuhe pro Jahr zu erwerben. Bei etwa jedem fünften (22,2%) sind es 4 Paare oder mehr. Die dabei stark präferierte Schuhart sind die in dieser Arbeit untersuchten Sneakers. 73,1% der Befragten geben diese als ihren Favoriten an, gefolgt von Sportschuhen (7,7%). Darüber hinaus sind Schuhe der Marke Adidas der Stichprobe sehr gut bekannt, 77,9% sind in Besitz von mindestens einem Paar (vgl. Tab. 27, Anhang S. A28).

Voraussetzung für die weitere Analyse ist eine Normalverteilung der Daten. Der zentrale Grenzwertsatz besagt, dass die Stichprobenverteilung des Mittelwerts für jede unabhängige Zufallsvariable normalverteilt (bzw. fast normalverteilt) sein wird, wenn die Stichprobengröße groß genug ist. Je größer die Stichprobe wird, desto näher wird die Stichprobenverteilung normalverteilt sein. Dank des zentralen Grenzwertsatzes können Hypothesentests durchgeführt werden, auch wenn die Grundgesamtheit nicht normalverteilt ist, vorausgesetzt, die Stichprobe ist ausreichend groß. Die meisten Statistikbücher geben als Empfehlung eine Stichprobengröße von $n > 30$ an (vgl. Polasek, 1997, S. 18), ab der von einer normalverteilten Stichprobenverteilung ausgegangen werden kann. In dieser Arbeit liegen die einzelnen Stichprobenumfänge über 30, weshalb Hypothesentests durchgeführt werden können.

4.2.2. Faktorenanalyse

Die Auswertung der Daten erfolgt mittels des Softwareprogramms SPSS Statistics. Im ersten Schritt dieser Analyse werden die Variablen auf Validität und Reliabilität geprüft. Hierzu wird eine Faktorenanalyse mit einer darauffolgenden Reliabilitätsanalyse angewendet. Die Faktorenanalyse ist ein Verfahren der Interdependenzanalyse zur Datenreduktion. Die Anwendung der konfirmatorischen Faktorenanalyse dient dazu, in einer Gruppe von Indikatoren eine Faktorstruktur zu ermitteln. Es wird also untersucht, ob die Indikatoren, die ein Konstrukt abbilden sollten, dies auch in den empirisch erhobenen Daten tun. Dadurch wird die Gültigkeit der verwendeten Itembatterien überprüft, um diese für die weiteren Analysen verwenden zu können (vgl. Backhaus, Erichson, Plinke & Weiber, 2016, S. 188).

Zur besseren Interpretation werden die sieben Dimensionen des RepTrak hier nicht zu einem Faktor verdichtet, sondern separat betrachtet. So können die verschiedenen Effekte der Stimuli noch genauer erkannt werden.

Vor der Anwendung der Faktorenanalyse sollten die Kaiser-Meyer-Olkin (KMO) Werte, welche die Korrelationen der Items miteinander angeben, betrachtet werden. Je näher das KMO-Kriterium dem Wert 1 kommt, desto mehr eignet sich die Gesamtheit aller Items für die Durchführung einer Faktorenanalyse. Bei einem Wert unter 0,5 sollte eine Faktorenanalyse nicht durchgeführt werden (vgl. Cleff, 2015, S. 220). Die KMO-Werte dieser Studie liegen bei allen Konstrukten über 0,5 (vgl. Tab. 4). Während sich bei der Reaktion auf einen Greenwashing-Skandal ein eher mäßiger Wert ergibt, weisen die KMO-Werte der RepTrak-Dimensionen, die Markenwahrnehmung (MW), das Markenverhalten (MV) sowie das Konstrukt Environmental Concern (EC) auf eine mittlere Korrelation der Items hin. Bei der Einstellung gegenüber der Marke sowie der Wahrnehmung von Greenwashing (WNGW) können sogar gute KMO-Werte verzeichnet werden (vgl. Kaiser & Rice, 1974, S. 112).

Anschließend wird der Bartlett-Test auf Sphärizität betrachtet. Dieser überprüft die Nullhypothese, ob die Korrelationsmatrix der beobachteten Variablen in der Grundgesamtheit gleich der Einheitsmatrix ist. Damit die Faktorenanalyse funktionieren kann, muss eine gewisse Beziehung

Tabelle 3: Darstellung der Stichprobe mit ihren Ausprägungen und Häufigkeiten

Quelle: Eigene Darstellung

Variable	Ausprägung	n	%
Alter	<25	277	53,6
	25-29	99	19,1
	30-39	43	8,3
	> 39	98	18,9
Geschlecht	weiblich	262	50,7
	männlich	255	49,3
Höchster Bildungsabschluss.	ohne Schulabschl.	32	6,2
	Haupt-/Realschule / mittlere Reife	64	12,4
	Abitur	222	42,9
	(Fach)Hochschulabschluss	187	36,2
	Promotion / Habilitation	7	1,4
	Keine Angabe	5	1,0
Beschäftigungsverhältnis	Ohne Arbeitsverh.	3	0,6
	SchülerIn	73	14,1
	StudentIn	192	37,1
	Auszubildende/r	25	4,8
	Angestellte/r	165	31,9
	BeamtIn	20	3,9
	Selbstständige/r	24	4,6
	Rentner/in	7	1,4
	Sonstiges	8	1,5
Haushaltsnettoeinkommen	< 500€	38	7,4
	500-1000	56	10,8
	1000-2000€	78	15,1
	2000-3000€	68	13,2
	3000-4000€	69	13,3
	> 4000	117	22,6
	Keine Angabe	91	17,6

zwischen einigen Variablen bzw. Gruppen von Variablen vorhanden sein. In dieser Studie kann die Nullhypothese des Bartlett-Tests bei allen Faktoren abgelehnt werden, sodass alle Korrelationen der Korrelationsmatrix größer als null sowie auf dem 1%-Niveau signifikant sind. Damit ist die Durchführung einer Faktorenanalyse möglich (vgl. Bühner, 2010, S. 207).

Im nächsten Schritt werden nach dem Kaiser-Kriterium alle Faktoren mit einem Eigenwert oberhalb von eins extrahiert und alle anderen Faktoren verworfen, um die durchschnittlich erfasste Varianz zu überprüfen. Der Eigenwert ist dabei als Summe der quadrierten Faktorladungen eines Faktors über alle Indikatoren definiert (vgl. Backhaus et al., 2016, S. 188). Der kumulierte Anteil der durch die Faktoren erklärten Varianz (DEV) liegt bei allen Konstrukten über 50% (vgl. Tab. 4). Dies erfüllt die Mindestanforderungen, sodass die Reliabilität des Konstruktes und der es abbildenden Indikatoren untersucht werden kann (vgl. Homburg & Baumgartner, 1995, S. 172; Homburg & Giering, 1996, S. 12). Das Konstrukt „Reaktion auf einen Greenwashing-Skandal“ wird dabei in zwei Faktoren aufgeteilt. Diese ergeben sich zum einen aus den Items 1 und 4 und zum anderen aus den Items 2 und 3 (vgl. Anhang S. A17-A18).

Die Faktorladungen geben die Größe der Bedeutung der Variablen für den Faktor an. Eine von vielen Forschern verwendete Faustregel ist, Items mit Ladungen von 0,7 oder mehr zu akzeptieren (vgl. Hulland, 1999, S. 198; Carmines & Zeller, 1979, S. 66). Mit Ausnahme der Dimension Rep. Innovation sind alle Faktorladungen größer als 0,7, sodass das Kriterium für ein valides Ergebnis erfüllt ist. Die Rep. Innovation besteht aus drei Items, die geringe Faktorladung (< 0,7) tritt dabei nur bei einem Item auf (Die Marke war die Erste in ihrer Branche). Da Adidas tatsächlich den Sportschuh revolutioniert hat und dies insbesondere in Deutschland weitestgehend bekannt ist, ist dieser Ausfall nicht sonderlich überraschend. Aufgrund des Schwellenwerts von Reitmeyer (2000, S. 96), welcher bei 0,4 für die Faktorladungen liegt, bleibt dieses Item im weiteren Verlauf der Analyse bestehen.

Mithilfe der Reliabilitätsanalyse wird nun die Zuverlässigkeit und Reproduzierbarkeit der Itembatterien überprüft. Dazu wird in dieser Studie Cronbachs Alpha verwendet. Cronbachs Alpha ist ein Maß für die interne Konsistenz der In-

Tabelle 4: Ergebnisse der Faktoren- und Reliabilitätsanalyse

Quelle: Eigene Darstellung

Konstrukt	Faktorladung	DEV	KMO	Cronbachs Alpha
Einstellung	0,745-0,855	64,239%	0,801*	0,860
Rep. Produkte	0,775-0,833	63,794%	0,789*	0,809
Rep. Innovation	0,637-0,822	57,885%	0,606*	0,623
Rep. Arbeitsumfeld	0,903-0,936	83,590%	0,737*	0,901
Rep. Leitung	0,863-0,902	77,184%	0,722*	0,852
Rep. Ges. Einbindung	0,843-0,858	72,090%	0,713*	0,803
Rep. Führung	0,740-0,796	57,904%	0,717*	0,756
Rep. Performance.	0,743-0,877	68,001%	0,653*	0,751
MW	0,819-0,888	72,319%	0,692*	0,803
MV	0,851-0,934	81,938%	0,708*	0,888
Reaktion Skandal Vertrauen	0,918-0,918	84,191%	0,500*	0,811
Reaktion Skandal Neubewertung	0,892-0.892	79,596%	0,500*	0,742
WNGW	0,711-0,877	64,426%	0,845*	0,888
EC	0,794-0,878	71,954%	0,690*	0,795

* $p < 0{,}01$

dikatoren eines Konstruktes (vgl. Peterson, 1994, S. 382). Der aus der Berechnung von Cronbachs Alpha resultierende Wertebereich liegt zwischen Null und Eins. Hohe Werte nahe Eins deuten grundsätzlich auf ein hohes Maß an Reliabilität der Messskala hin. Da eine inferenzstatistische Beurteilung von Cronbachs Alpha nicht möglich ist, werden in der Literatur bezüglich eines Mindestwertes für eine ausreichende Reliabilität verschiedene Werte genannt. Cronbach schreibt, dass hohe Werte wünschenswert seien, aber eine Skala nicht perfekt sein müsse, um verwendbar zu sein (vgl. Cronbach, 1951, S. 332). Allgemein gefordert wird ein Mindestwert von 0,6 (vgl. Malhotra & Dash, 2016, S. 308). Demnach weisen die Alphas bei den RepTrak-Dimensionen Innovation, Führung und Performance sowie das Konstrukt Environmental Concern auf eine akzeptable Reliabilität hin. Die anderen Dimensionen und Konstrukte erreichen Werte über 0,8 bzw. sogar teils über 0,9, was gute bis exzellente Reliabilität beweist (vgl. George/Mallery 2003, S. 231).

Zusammengefasst kann durch die Faktorenanalyse festgestellt werden, dass die Items der jeweiligen Konstrukte eine hohe Korrelation aufweisen. Somit konnte sowohl die Validität als auch die Reliabilität der verwendeten Items bzw. Konstrukte nachgewiesen werden (vgl. Bühner, 2010, S. 180). Insgesamt können also auf Basis der ausgewählten Items und gebildeten Konstrukte weitere Analysen durchgeführt werden, wobei jedes Konstrukt auch je einen Faktor darstellt. Lediglich das Konstrukt „Reaktion auf einen Greenwashing-Skandal“ wird im Folgenden anhand zweier Faktoren abgebildet (vgl. Tab. 4).

4.2.3. Varianzanalyse

Im Folgenden werden die Mittelwertvergleiche dieser Studie betrachtet. Mittels einer Varianzanalyse (eng.: analysis of variance; Anova) können aus den beobachteten Mittelwerten mehrerer Stichproben Rückschlüsse auf die Mittelwerte der Grundgesamtheit gezogen werden. Dazu wird die zu testende Hypothese als Nullhypothese H_0 definiert. Diese besagt, dass die Mittelwerte der Grundgesamtheiten gleich sind, also zwischen den Gruppen keine Unterschiede bestehen. H_0 wird bei einer Irrtumswahrscheinlichkeit (Signifikanz) von $p < 0{,}05$ verworfen. Das bedeutet, dass die Wahrscheinlichkeit des irrtümlichen Zurückweisens der Nullhypothese so gering ist, dass praktisch kein Fehler begangen werden kann. Bei Ablehnung der Nullhypothese greift die Alternativhypothese H_1, welche besagt, dass sich die Mittelwerte der Gruppen in der Grundgesamtheit voneinander unterscheiden. Eine Anova kann allerdings lediglich feststellen, ob in einer Gruppe von Mittelwerten signifikante Unterschiede bestehen. Werden drei oder mehr Gruppen verglichen, wird daher, sofern die Nullhypothese verworfen wird, anschließend ein Post-hoc Test durchgeführt. Post-hoc Tests mit paarweisen Mittelwertvergleichen geben Auskunft darüber, welche Mittelwerte genau sich signifikant voneinander unterscheiden (vgl. Ruxton & Beauchamp, 2008, S. 690). Für die Entscheidung, welcher Post-hoc Test verwendet werden sollte, muss zunächst die Varianzhomogenität überprüft werden. Der Levene-Test prüft die Gleichheit der Varianzen zweier Grundgesamtheiten. Ist die Signifikanz für den Levene-Test höher als 0,05 wird die Annahme gleicher Varianzen nicht zurückgewiesen und Varianzhomogenität angenommen (vgl. Brosius, 2006, S. 478-483; Tomarken & Serlin, 1986, S. 93). In dieser Untersuchung kann Varianzhomogenität auch bei Werten unter 0,05 angenommen werden, da die Stichprobenumfänge der einzelnen Gruppen ähnlich sind (vgl. Bortz, 2005, S. 133). Daher wird durchgängig der Scheffe-Test als Post-hoc Test verwendet. Die Ergebnisse der

Levene-Tests sind in Tab. 28 (vgl. Anhang S. A29) aufgeführt.

Zu Beginn der Mittelwertvergleiche wird die Einstellung gegenüber der Marke Adidas anhand von fünf Items analysiert. Insgesamt, also über alle Gruppen hinweg, ist diese mit einem Mittelwert von 3,65 relativ hoch. Die Marke Adidas genießt demnach auch bei den Probanden dieser Studie ein hohes Ansehen. Um auszuschließen, dass in Bezug auf die Einstellung signifikante Unterschiede zwischen den Gruppen bestehen, welche die Untersuchungsergebnisse verfälschen könnten, wird nun eine Anova durchgeführt. Der p-Wert liegt hier mit 0,309 deutlich über 0,05, weshalb die Nullhypothese nicht verworfen werden kann. Es besteht demnach kein signifikanter Unterschied zwischen den einzelnen Gruppen (vgl. Tab. 29, Anhang S. A29). Nachdem die Probanden die Fragen zu der Einstellung gegenüber Adidas beantwortet haben, wurde ihnen, je nachdem welcher Gruppe sie zugeordnet sind, eine Werbeanzeige für neue Adidas Sneakers gezeigt (vgl. Anhang S. A4-A8). Im Folgenden werden die Effekte dieser unterschiedlichen CSR-Initiativen, Greenwashing-Methoden sowie Greenwashing-Skandale auf die Reputation von Adidas analysiert. Dazu werden die verschiedenen Stichproben zunächst einzeln betrachtet. So kann herausgefunden werden, ob unterschiedliche Initiativen bzw. Methoden unterschiedliche Effekte auf die Reputation der Marke Adidas haben.

Die Tabellen 5 und 6 zeigen die Mittelwerte der Konstrukte für alle sieben Gruppen sowie den Mittelwert über alle Gruppen hinweg (Spalte „Gesamt“) und den p-Wert, also die Signifikanz. Da die p-Werte bei allen Faktoren unter 0,05 liegen, kann die Nullhypothese, dass keine Gruppenunterschiede bestehen, abgelehnt werden. Das bedeutet, dass signifikante Unterschiede zwischen den Gruppen existieren. Um herauszufinden, zwischen welchen Gruppenpaaren diese Unterschiede tatsächlich bestehen, wird daraufhin der Post-hoc Test durchgeführt.

Für den Faktor Rep. Produkte lassen sich dabei Unterschiede der Gruppe 7 zu allen anderen Gruppen außer der Gruppe 4, also der zweiten Greenwashing-Skandal-Gruppe feststellen ($p < 0,01$). Der Mittelwert liegt mit 3,28 deutlich unter denen der anderen Gruppen, der Greenwashing-Skandal in Bezug auf vage Formulierungen hat also zu einer geringeren Produktreputation geführt.

Bei den RepTrak-Faktoren Innovation, Leitung sowie gesellschaftliche Einbindung zeigen sich weiterhin signifikante Unterschiede zwischen den beiden Skandal-Gruppen (4 & 7) und den CSR-Initiativen (3 & 6). Während die CSR-Initiativen auf ähnliche Werte wie die Kontrollgruppe kommen, liegen die Mittelwerte der Skandal-Gruppen auch hier signifikant niedriger ($p < 0,05$). Ähnliche Ergebnisse zeigen sich bei den Faktoren Markenwahrnehmung und Markenverhalten.

Darüber hinaus können auch bei den weiteren Faktoren Tendenzen dafür erkannt werden, dass die Werbeanzeigen der Kontrollgruppe sowie der CSR-Initiativen und Greenwashing-Methoden ähnliche Effekte auf die Reputation der Marke Adidas haben, während bei den Greenwashing-Skandalen der Gruppen 4 und 7 negative Effekte auftreten. Diese Unterschiede sind allerdings nicht auf das 5%-Niveau signifikant.

Um die Unterschiede der verschiedenen (vermeintlichen) Green Marketing-Aktivitäten besser erkennen zu können, werden nachfolgend die Gruppen 2 & 5, 3 & 6 sowie 4 & 7 zusammen betrachtet. Das heißt, es wird nicht mehr nach der jeweiligen Initiative bzw. Methode unterschieden, sondern ausschließlich zwischen CSR, Greenwashing sowie Greenwashing-Skandal.

Auch hier liegen die p-Werte der Anova allesamt unter 0,05, es existieren also signifikante Gruppenunterschiede und der Post-hoc Test kann durchgeführt werden (vgl. Tab. 7 & 8). Dieser bestätigt die Tendenzen aus der vorherigen Analyse. Bei fast allen RepTrak-Faktoren zeigen sich dieselben Ergebnisse: Zwischen der Kontrollgruppe, den CSR-Initiativen und Greenwashing-Methoden bestehen keine signifikanten Unterschiede, dafür aber zwischen den Greenwashing-Skandalen und allen anderen Gruppen ($p < 0,05$). Nur bei den Faktoren Rep. Arbeitsumfeld und Rep. Performance bestehen diese Unterschiede lediglich zwischen den Skandalen und den CSR-Initiativen.

Diese Ergebnisse machen sich durch die deutlich geringeren Mittelwerte bei den Gruppen mit den Greenwashing-Skandalen bemerkbar. Insbesondere bei den RepTrak-Faktoren Leitung sowie gesellschaftliche Einbindung, aber auch bei der Rep. Führung haben die Probanden aus den Skandal-Gruppen die Reputation erheblich schlechter bewertet als die der anderen Gruppen (vgl. Tab 7). Dem nachgelagert lassen sich diese Unterschiede auch bei den Faktoren des BPI, also der Markenwahrnehmung und dem Markenverhalten erkennen (vgl. Tab 8). Bei beiden Faktoren gibt es signifikante Unterschiede zwischen den Greenwashing-Skandalen und den anderen drei Gruppen ($p < 0,05$). Auch die Preisakzeptanz ist mit durchschnittlich 62,55€ bei den Greenwashing-Skandalen deutlich niedriger. Signifikante Unterschiede können hier allerdings nur zwischen den Skandalen und den CSR-Initiativen festgestellt werden.

Im nächsten Schritt wird die Wahrnehmung von Greenwashing (WNGW) untersucht. Dazu werden zunächst die Unterschiede zwischen den Gruppen betrachtet und im Anschluss analysiert, inwieweit die Wahrnehmung die Effekte der CSR-Initiativen und Greenwashing-Methoden auf die Reputation beeinflusst. Die Tabelle 9 zeigt, dass Greenwashing ein wenig stärker von den Probanden wahrgenommen wurde, denen auch tatsächlich eine Greenwashing-Methode gezeigt wurde. Am höchsten ist der Mittelwert bei Gruppe 5, also der Greenwashing-Methode Vagheit. Im Post-hoc Test können allerdings keine signifikanten Unterschiede festgestellt werden.

Ein wenig andere Ergebnisse zeigen sich, wenn erneut die Gruppen 3 und 6 sowie 2 und 5 gemeinsam betrachtet werden. Zwar ist der Wert bei den Greenwashing-Methoden nur etwas höher (Ø 3,56), als bei den CSR-Initiativen (Ø 3,34), hier sind die Unterschiede jedoch signifikant ($p < 0,05$) (vgl. Tab. 10).

Nun wird analysiert, inwieweit die Wahrnehmung von Greenwashing die Effekte der CSR-Initiativen und Green-

Tabelle 5: Mittelwertvergleich der Reputation über alle Gruppen

Quelle: Eigene Darstellung

Faktor	Gruppe							Gesamt	p-Wert
	KG	CSR		GW		Skandal			
	1	3	6	2	5	4	7		
Rep. Produkte	3,74	3,89	3,84	3,84	3,77	3,46	3,28	3,70	0,000
Rep. Innov.	3,49	3,72	3,57	3,54	3,41	3,09	3,15	3,43	0,000
Rep. Arb. Umf.	2,76	2,97	2,95	2,80	2,92	2,70	2,62	2,82	0,024
Rep. Leitung	2,94	3,21	2,99	2,84	2,97	2,68	2,49	2,88	0,000
Rep. Ges. Einb.	2,99	3,25	3,04	2,97	3,01	2,59	2,54	2,92	0,000
Rep. Führung	3,40	3,50	3,44	3,50	3,40	3,19	3,15	3,37	0,000
Rep. Perf.	4,13	4,19	4,24	4,11	4,01	3,98	3,89	4,08	0,011

Tabelle 6: Mittelwertvergleich MW und MV über alle Gruppen

Quelle: Eigene Darstellung

Faktor	Gruppe							Gesamt	p-Wert
	KG	CSR		GW		Skandal			
	1	3	6	2	5	4	7		
MW	3,60	3,88	3,61	3,66	3,52	3,18	3,21	3,53	0,000
MV	3,58	3,81	3,54	3,66	3,50	2,92	3,25	3,47	0,000
Preisakzeptanz (€)	70,75	88,32	85,74	77,83	73,96	61,56	63,41	74,68	0,000

Tabelle 7: Mittelwertvergleich der Reputation mit zusammengefassten Gruppen

Quelle: Eigene Darstellung

Faktor	Gruppe				p-Wert
	Kontrollgruppe	CSR- Initiativen	GW- Methoden	GW- Skandale	
Rep. Produkte	3,74	3,87	3,80	3,37	0,000
Rep. Innovation	3,49	3,64	3,48	3,12	0,000
Rep. Arb. Umf.	2,76	2,96	2,86	2,65	0,004
Rep. Leitung	2,94	3,10	2,91	2,58	0,000
Rep. Ges. Einb.	2,99	3,14	2,99	2,57	0,000
Rep. Führung	3,40	3,47	3,45	3,17	0,000
Rep. Performance	4,13	4,21	4,06	3,93	0,002

Tabelle 8: Mittelwertvergleich MW und MV mit zusammengefassten Gruppen

Quelle: Eigene Darstellung

Faktor	Gruppe				p-Wert
	Kontroll- gruppe	CSR- Initiativen	GW- Methoden	GW- Skandale	
MW	3,60	3,74	3,60	3,20	0,000
MV	3,58	3,67	3,58	3,09	0,000
Preisakzeptanz (€)	70,75	86,99	75,95	62,55	0,000

washing-Methoden auf die Reputation beeinflusst. Dazu fungiert der Faktor „Wahrnehmung von Greenwashing“ als unabhängige Variable und die Faktoren des RepTrak sowie die Markenwahrnehmung und das Markenverhalten als abhängige. In der Spalte „WNGW gering“ sind die Mittelwerte der Probanden gelistet, die bei der Wahrnehmung von Greenwashing einen Mittelwert von bis zu 3,0 angegeben haben. Liegt der Mittelwert darüber, fließen die Daten in die Mittelwerte

Tabelle 9: Mittelwertvergleich der WNGW über alle Gruppen

Quelle: Eigene Darstellung

Faktor	**Gruppe**				**p-Wert**
	CSR-Initiativen		GW-Methoden		
	Gruppe 3	Gruppe 6	Gruppe 2	Gruppe 5	
Wahrnehmung von GW	3,27	3,40	3,50	3,63	0,048

Tabelle 10: Mittelwertvergleich der WNGW mit zusammengefassten Gruppen

Quelle: Eigene Darstellung

Faktor	**Gruppe**		**p-Wert**
	CSR-Initiativen	GW-Methoden	
Wahrnehmung von GW	3,34	3,56	0,013

der Spalte „WNGW stark“ ein. Zum Vergleich werden auch noch die Mittelwerte aus der Kontrollgruppe angegeben (vgl. Tab. 11 & 12).

Tabelle 11 zeigt die Werte für die CSR-Initiativen, also die Gruppen 3 und 6. Von den 148 Probanden aus den beiden Gruppen haben 107 Greenwashing eher gering wahrgenommen und 41 fälschlicherweise stark. Abgesehen von dem Faktor Rep. Performance sind die Unterschiede überall auf das 1%-Niveau signifikant. Dabei zeigen sich große Differenzen zwischen den Mittelwerten. Haben die Probanden Greenwashing stark erkannt (obwohl es sich hier um CSR-Initiativen handelt), sind die RepTrak-Faktoren, die Markenwahrnehmung und das Markenverhalten erheblich schlechter bewerte worden. Die Werte der Kontrollgruppe liegen dabei stets zwischen den anderen beiden (vgl. Tab. 11).

Bei den Gruppen 2 und 5, also den Greenwashing-Methoden, sind die Ergebnisse ähnlich. Hier haben 103 Probanden Greenwashing eher gering wahrgenommen und 49 stark. Im Vergleich zu den 27,7%, der Probanden, die bei den CSR-Initiativen Greenwashing stark wahrgenommen haben, sind es hier mit 32,2% folglich etwas mehr. Die Unterschiede sind mit Ausnahme der RepTrak-Faktoren Innovation und Performance bei allen Faktoren signifikant ($p < 0{,}01$). Wenn Greenwashing (hier fälschlicherweise) nicht erkannt wurde, so zeigen sich bessere Bewertungen bei den RepTrak-Faktoren, der Markenwahrnehmung sowie dem Markenverhalten. Tendenziell sind die Bewertungen sogar etwas besser als die der Kontrollgruppe (vgl. Tab. 12).

Im Folgenden werden die wahrgenommenen Motivationen für die (vermeintlich) nachhaltigen Unternehmenspraktiken der Marke Adidas betrachtet. Den Probanden, denen CSR-Initiativen oder Greenwashing-Methoden gezeigt wurden, konnten anhand je einer Frage angeben, ob sie der Meinung sind, die Motivation für die nachhaltige Markenausrichtung sei strategisch/wirtschaftlich bzw. ökologisch/sozial. Zwischen den vier Gruppen können dabei keine signifikanten Unterschiede festgestellt werden. (vgl. Tab. 30, Anhang S. A30).

Darüber hinaus wird noch untersucht, welche Motivation die Probanden hinter den tatsächlichen CSR-Initiativen vermuten und welche Auswirkungen dies hat. Die Tabellen 13 und 14 geben Auskunft über die wahrgenommene strategische sowie ökologische Motivation. In Bezug auf die Wahrnehmung der strategischen Motivation lassen sich dabei kaum signifikante Unterschiede erkennen. Lediglich die Rep. Führung wird leicht besser bewertet, wenn eine starke strategische Motivation wahrgenommen wird ($p < 0{,}05$). Bei den Faktoren Markenwahrnehmung und Markenverhalten lassen sich lediglich Tendenzen erkennen, dass eine strategische Motivation zu höheren Werten führt (vgl. Tab. 13).

Insgesamt haben lediglich 39 Probanden eine geringe strategische Motivation wahrgenommen; 109 Probanden vermuten eine starke strategische Motivation hinter den CSR-Initiativen (vgl. Tab. 13).

Weitaus stärkere, signifikante, Unterschiede ergeben sich bei der Analyse der wahrgenommenen ökologischen Motivation. Mit Ausnahme des Faktors Rep. Performance ist der Mittelwert bei allen Faktoren erheblich höher, sofern die Probanden eine starke ökologische Motivation hinter dem Green Marketing annehmen. Zwar vermuten nur etwa ein Drittel diese starke ökologische Motivation, diese Probanden bewerten die Marke Adidas dafür erheblich besser. Insbesondere bei den RepTrak-Faktoren Arbeitsumfeld, Leitung und gesellschaftliche Einbindung sowie dem Markenverhalten lassen sich starke Unterschiede erkennen (vgl. Tab. 14).

Eine weitere Analyse befasst sich damit, ob die Höhe des Umweltinvolvements einen Einfluss darauf hat, ob die Probanden die CSR-Initiativen und Greenwashing-Methoden als solche erkennen, oder falsch beurteilen. Dazu werden sowohl das Umweltbewusstsein (EC), als auch das Umweltwissen (EK) betrachtet. Von den Probanden aus den CSR-Gruppen haben 39 ein eher geringes und 109 ein hohes Umweltbewusstsein. Die Mittelwerte der beiden Gruppen unterscheiden sich dabei allerdings nicht signifikant, der p-Wert liegt bei 0,728 (vgl. Tab. 15).

Andere Ergebnisse zeigen sich bei der Betrachtung des

Tabelle 11: Mittelwertvergleich je nach Stärke der WNGW bei den CSR-Initiativen

Quelle: Eigene Darstellung

Faktor	WNGW gering	WNGW stark	Kontrollgruppe	p-Wert
	n = 107	n = 41		
Rep. Produkte	3,96	3,62	3,74	0,005
Rep- Innovation	3,83	3,17	3,49	0,000
Rep. Arbeitsumfeld	3,11	2,57	2,76	0,000
Rep. Leitung	3,18	2,62	2,94	0,000
Rep. Ges. Einb.	3,35	2,62	2,97	0,000
Rep. Führung	3,56	3,24	3,40	0,000
Rep. Performance	4,25	4,12	4,13	0,183
MW	3,85	3,46	3,60	0,008
MV	3,83	3,27	3,57	0,004

Tabelle 12: Mittelwertvergleich je nach Stärke der WNGW bei den GW-Methoden

Quelle: Eigene Darstellung

Faktor	WNGW gering	WNGW stark	Kontrollgruppe	p-Wert
	n = 103	n = 49		
Rep. Produkte	3,93	3,54	3,74	0,000
Rep. Innovation	3,50	3,42	3,49	0,464
Rep. Arbeitsumfeld	2,98	2,62	2,76	0,005
Rep. Leitung	3,11	2,48	2,94	0,000
Rep. Ges. Einb.	3,19	2,58	2,97	0,000
Rep. Führung	3,55	3,24	3,40	0,001
Rep. Performance	4,11	3,97	4,13	0,210
MW	3,78	3,20	3,60	0,000
MV	3,79	3,16	3,57	0,000

Tabelle 13: Mittelwertvergleich je nach Stärke der strategischen Motivation

Quelle: Eigene Darstellung

Faktor	Geringe strat. Motivation	Starke strat. Motivation.	p-Wert
	n = 39	n = 109	
Rep. Produkte	3,68	3,93	0,039
Rep. Innovation	3,63	3,65	0,906
Rep. Arbeitsumfeld	3,00	2,94	0,672
Rep. Leitung	3,11	3,09	0,902
Rep. Ges. Einb.	3,18	3,13	0,762
Rep. Führung	3,32	3,53	0,030
Rep. Performance	4,12	4,25	0,187
MW	3,64	3,78	0,378
MV	3,48	3,74	0,184

Umweltwissens. Denjenigen Probanden, die alle vier in dem Experiment gezeigten (vermeintlichen) Umweltlabel als zertifiziert bzw. gefälscht erkennen konnten, wird hier ein hohes Umweltwissen zugeschrieben. Während diese bei den CSR-Initiativen lediglich einen Mittelwert von 2,85 bei der Wahrnehmung von Greenwashing erreichen, liegt der Wert bei den anderen Probanden bei 3,43 ($p < 0,01$) (vgl. Tab. 16).

Die Probanden, die eine Greenwashing-Methode als Werbeanzeige vorgelegt bekommen haben, haben Greenwashing auch tatsächlich etwas stärker wahrgenommen (vgl. Tab. 9). Die Höhe des Umweltbewusstseins hat darauf allerdings ebenfalls keinen signifikanten Einfluss. Der p-Wert liegt mit

Tabelle 14: Mittelwertvergleich je nach Stärke der ökologischen Motivation

Quelle: Eigene Darstellung

Faktor	Geringe ökol. Motivation	Starke ökol. Motivation.	p-Wert
	n = 99	n = 49	
Rep. Produkte	3,69	4,22	0,000
Rep. Innovation	3,45	4,03	0,000
Rep. Arbeitsumfeld	2,73	3,42	0,000
Rep. Leitung	2,76	3,78	0,000
Rep. Ges. Einb.	2,82	3,79	0,000
Rep. Führung	3,33	3,77	0,000
Rep. Performance	4,17	4,31	0,103
MW	3,49	4,24	0,000
MV	3,41	4,20	0,000

Tabelle 15: Mittelwertvergleich je nach Stärke des EC bei den CSR-Initiativen

Quelle: Eigene Darstellung

Faktor	Geringes EC	Hohes EC	p-Wert
	n = 39	n = 109	
Wahrnehmung von GW	3,29	3,35	0,728

Tabelle 16: Mittelwertvergleich je nach Stärke des EK bei den CSR-Initiativen

Quelle: Eigene Darstellung

Faktor	Geringes EK	Hohes EK	p-Wert
	n = 124	n = 24	
Wahrnehmung von GW	3,43	2,85	0,002

0,116 über 0,05 (vgl. Tab. 17).

Im Gegensatz dazu zeigen sich signifikante Unterschiede, wenn das Umweltwissen untersucht wird. Probanden mit einem hohen Umweltwissen haben Greenwashing hier etwas stärker wahrnehmen können. Die Ergebnisse sind auf das 5%-Niveau signifikant (vgl. Tabelle 18).

Ähnlich zu der vorigen Analyse wird noch der Einfluss des Alters auf die Wahrnehmung von Greenwashing untersucht. Dazu werden die Probanden in zwei Gruppen aufgeteilt: Der Gruppe „Junges Alter“ sind alle Probanden bis 25 Jahre und der Gruppe „Hohes Alter“ alle Probanden ab 26 Jahre zugeordnet. Innerhalb der CSR-Gruppen haben die jüngeren Probanden Greenwashing zwar etwas stärker wahrgenommen, der p-Wert liegt allerdings bei 0,367, der Unterschied ist somit nicht signifikant (vgl. Tab. 19).

Auch in den Greenwashing-Gruppen haben die jüngeren Probanden eher irreführende bzw. falsche Behauptungen wahrgenommen als die älteren. Mit einem Mittelwert von 3,71 liegt diese Wahrnehmung auch deutlich über der der älteren Probanden (Ø 3,33). Diese Ergebnisse sind zudem signifikant ($p < 0{,}001$) (vgl. Tab. 20).

Als letzte Analyse von möglichen Einflüssen der Konsumenteneigenschaften auf die Wahrnehmung von Greenwashing wird das Geschlecht betrachtet. Die Tabellen 31 und 32 (vgl. Anhang S. A30) zeigen jeweils die Wahrnehmung der weiblichen und männlichen Probanden sowohl aus den CSR-Gruppen als auch aus den Greenwashing-Gruppen. Da der Mittelwert der Frauen innerhalb der CSR-Gruppen niedriger ist und innerhalb der Greenwashing-Gruppen höher, zeigt die Tendenz, dass Frauen CSR-Initiativen und Greenwashing-Methoden eher als solche erkennen. Die p-Werte sind allerdings bei beiden Varianzanalysen über 0,05, die Unterschiede sind also nicht signifikant.

Abschließend werden noch die Reaktionen auf einen Greenwashing-Skandal analysiert. Diese wurde bei den Probanden der Gruppen 4 und 7, also jenen, denen ein Zeitungsartikel über einen Greenwashing-Skandal bei Adidas gezeigt wurde, überprüft. Mittels der Faktorenanalyse ist das Konstrukt in zwei Faktoren aufgeteilt worden. Der erste bildet den Vertrauensverlust in die Marke ab und der zweite, ob die Probanden die Möglichkeit einer Neubewertung in Betracht ziehen, oder mit der Marke gänzlich abgeschlossen haben. Je höher der Mittelwert ist, desto größer ist der Vertrauensverlust bzw. desto weniger können sich der Pro-

Tabelle 17: Mittelwertvergleich je nach Stärke des EC bei den GW-Methoden

Quelle: Eigene Darstellung

Faktor	Geringes EC	Hohes EC	p-Wert
	n = 39	n = 113	
Wahrnehmung von GW	3,41	3,62	0,116

Tabelle 18: Mittelwertvergleich je nach Stärke des EK bei den GW-Methoden

Quelle: Eigene Darstellung

Faktor	Geringes EK	Hohes EK	p-Wert
	n = 111	n = 41	
Wahrnehmung von GW	3,50	3,73	0,028

Tabelle 19: Mittelwertvergleich je nach Alter der Probanden bei den CSR-Initiativen

Quelle: Eigene Darstellung

Faktor	Junges Alter	Hohes Alter	p-Wert
	n = 89	n = 59	
Wahrnehmung von GW	3,39	3,26	0,367

Tabelle 20: Mittelwertvergleich je nach Alter der Probanden bei den GW-Methoden

Quelle: Eigene Darstellung

Faktor	Junges Alter	Hohes Alter	p-Wert
	n = 94	n = 58	
Wahrnehmung von GW	3,71	3,33	0,001

banden eine Neubewertung vorstellen. Während die Werte bei dem Faktor Neubewertung nahezu identisch sind, ist der Vertrauensverlust bei der Gruppe 4, also der Greenwashing-Methode Vagheit deutlich größer. Der Mittelwert beträgt hier 3,12 im Vergleich zu 2,73 bei der Gruppe 7 ($p <. 0,01$) (vgl. Tab. 21).

In diesem Zuge wird noch untersucht, ob das Geschlecht einen Einfluss darauf hat, wie stark die Reaktion auf einen Greenwashing-Skandal ausfällt. Wie anhand der Tabelle 33 (vgl. Anhang S. A31) zu erkennen, ist dies nicht der Fall. Die Mittelwerte zwischen den Geschlechtern sind sehr ähnlich und die Unterschiede nicht signifikant.

4.2.4. Mess- und Strukturmodell

Aus den Ergebnissen der Varianzanalysen im letzten Teil wird bereits deutlich, dass die Wahrnehmung von Greenwashing eine wesentliche Rolle bei den Effekten von CSR-Initiativen und Greenwashing-Methoden auf die Reputation sowie die Markenwahrnehmung und das Markenverhalten spielt. Aufgrund dessen werden im Folgenden die Kausalbeziehungen dieser Faktoren mithilfe eines Mess- und Strukturmodells untersucht. Das Ziel einer Kausalanalyse ist die Verifizierung der Ursache-Wirkungs-Beziehung zwischen Variablen. Empirisch gemessene Varianzen und Kovarianzen zwischen Indikatorvariablen lassen Rückschlüsse auf Abhängigkeitsbeziehungen zwischen latenten Variablen (Konstrukten) zu. Durch die strukturelle Gleichungsmodellierung können komplexe Effektstrukturen zwischen verschiedenen unabhängigen und abhängigen Konstrukten dargestellt und verifiziert werden. Als Analysetool für dieses Variance-Based Structural Equation Model (PLS-SEM) wird die Software SmartPLS verwendet. Anhand des Messmodells werden dazu zunächst die Zusammenhänge zwischen den Konstrukten und ihren jeweiligen Indikatorvariablen untersucht, um die Reliabilität und Validität der Konstrukte zu verifizieren. In diesem Fall ist das Messmodell reflektiv, da die Indikatoren von dem latenten Konstrukt verursacht werden. Die entsprechenden Werte sind in der Tabelle 22 aufgeführt.

Zur Untersuchung der internen Konsistenz der Konstrukte wird die Composite Reliability (CR) betrachtet. Diese misst die Korrelation zwischen Items und Konstrukt und spiegelt somit wider, ob ein Konstrukt geeignet ist, um seine Indikatoren zu erklären. Die CR sollte für jedes Konstrukt größer als 0,6 sein, um interne Konsistenz gewährleisten zu können

Tabelle 21: Mittelwertvergleich der Reaktion auf einen Skandal

Quelle: Eigene Darstellung

Faktor	Gruppe 4	Gruppe 7	p-Wert
Reaktion Skandal Vertrauen	3,12	2,73	0,008
Reaktion Skandal Neubewertung	2,70	2,69	0,926

(vgl. Bagozzi & Yi, 1988, S. 80). In dieser Studie weisen alle Konstrukte Werte über dieser Grenze auf (vgl. Tab. 22). Ein weiteres Maß zur Überprüfung der Reliabilität der Konstrukte ist Cronbachs Alpha. Auch dieses sollte Werte über 0,6 annehmen (vgl. Malhotra & Dash, 2016, S. 308), was hier ebenfalls erfüllt wird (vgl. Tab. 22).

Weiterhin werden bei latenten Konstrukten die Faktorladungen untersucht, welche auf die gleiche Weise wie die Ladungen in einer Faktorenanalyse interpretiert werden können. Wie bei der Faktorenanalyse weisen auch hier, außer bei der RepTrak-Dimension Innovation, alle Items Faktorladungen von mindestens 0,7 (vgl. Hulland, 1999, S. 198) auf (vgl. Tab. 22). Aufgrund der von **?**, S. 96 festgelegten Grenze von 0,4 wird aber dennoch auch mit allen Items dieses Konstrukts fortgefahren.

Darüber hinaus gilt ein Modell als konvergent, wenn die durchschnittliche erfasste Varianz (DEV) größer als 0,5 ist, was bedeutet, dass 50 Prozent oder mehr Varianz der Indikatoren erklärt werden (vgl. Homburg & Baumgartner, 1995, S. 172; Fornell & Larcker, 1981, S. 46). Auch dieses Kriterium wird hier von allen Konstrukten erfüllt (vgl. Tab. 22).

Die diskriminante Validität gibt das Ausmaß an, in dem sich ein gegebenes Konstrukt empirisch von anderen latenten Variablen im Strukturmodell unterscheidet. Zur Überprüfung dieser wird das Fornell-Larcker-Kriterium betrachtet. Demnach soll die DEV jedes latenten Konstrukts höher sein als die höchste quadrierte Korrelation des Konstrukts mit jedem anderen latenten Konstrukt. Das Fornell-Larcker-Kriterium wird ebenfalls erfüllt (vgl. Tab. 22). Auch das Kriterium der Kreuzladungen ist erfüllt. Jedes Item lädt auf sein eigenes Konstrukt höher als auf alle anderen Konstrukte (vgl. Chin et al., 1998, S. 321).

Nachdem nun das Messmodell analysiert und die interne Zusammensetzung der Konstrukte überprüft wurde, folgt nun das Strukturmodell. Hier werden die Beziehungen zwischen den Konstrukten untersucht, um Aufschlüsse über die Zusammenhänge und Kausalitätsbeziehungen geben zu können.

Bevor die strukturellen Beziehungen bewertet werden, muss die Multikollinearität untersucht werden, um sicherzustellen, dass sie die Regressionsergebnisse nicht verzerrt. Dazu wird der Varianzinflationsfaktor (VIF) der unabhängigen Variablen betrachtet. VIF-Werte von 5 oder mehr weisen auf Kollinearitätsprobleme zwischen den Konstrukten hin. Idealerweise sollten sie in der Nähe von 3 und darunter liegen (vgl. Akinwande, Dikko, Samson et al., 2015, S. 756). In diesem Modell liegen die VIF-Werte bei allen Konstrukten unter 3, sodass eine Verzerrung der Ergebnisse ausgeschlossen und mit der Analyse der Strukturmodells fortgefahren werden kann (vgl. Tab. 35, Anhang S. A33).

Mithilfe der Pfadkoeffizienten können im nächsten Schritt die Beziehungen zwischen zwei Konstrukten beschrieben werden. Dabei wird sowohl die Signifikanz als auch die Größe des Werts miteinbezogen. Pfadkoeffizienten fallen typischerweise in den Bereich von -1 bis +1. Bei Werten über 0,1 (absoluter Wert) kann von einer nennenswerten Beziehung gesprochen werden (vgl. Lohmöller, 1989, S. 60-61). Zunächst einmal zeigen sich erwartungsgemäß, abgesehen von dem Faktor Rep. Performance, durchweg moderate ($> 0{,}2$) bis relativ starke ($> 0{,}3$) negative Einflüsse von der Wahrnehmung von Greenwashing (WNGW) auf alle RepTrak-Faktoren. Die Werte sind auf das 1%-Niveau signifikant (vgl. Abb. 6). Auf die Markenwahrnehmung sowie das Markenverhalten bestehen keine nennenswerten Einflüsse. Bei den Einflüssen der RepTrak-Faktoren fallen insbesondere die des Faktors Rep. Produkte auf die Markenwahrnehmung (+ 0,231) sowie auf das Markenverhalten (+ 0,197) und der Einfluss des Faktors Rep. gesellschaftliche Einbindung auf die Markenwahrnehmung (+ 0,274) auf. Auch hier liegt der p-Wert unter 0,01. Darüber hinaus lassen sich geringe Einflüsse der RepTrak-Faktoren Leitung, Führung und Performance auf die Markenwahrnehmung erkennen ($p < 0{,}05$). Abschließend besteht noch ein starker ($> 0{,}5$; $p < 0{,}01$) positiver Einfluss der Markenwahrnehmung auf das Markenverhalten (vgl. Abb. 6). Tabelle 34 (vgl. Anhang S. A31-A32) gibt einen Überblick über alle Pfadkoeffizienten.

Die T-Statistik, bei der die Schwellenwerte bei 1,65 für das 10%-, 1,96 für das 5%- und 2,58 für das 1%-Signifikanzniveau liegen (vgl. Hair, Ringle & Sarstedt, 2011, S. 145), bestätigen diese Beziehungen zwischen den Konstrukten (vgl. Tab. 34, Anhang S. A31-A32).

Zur Überprüfung der Aussagekraft und Relevanz des Strukturmodells werden nun noch das Bestimmtheitsmaß R^2 sowie Stone-Geisser's Q^2-Wert untersucht. Das Bestimmtheitsmaß R^2 wird im Zusammenhang mit statistischen Modellen verwendet, deren Hauptzweck die Vorhersage zukünftiger Ergebnisse ist. Es misst die Varianz, die in den abhängigen Konstrukten erklärt wird und reicht von 0 bis 1, wobei höhere Werte eine größere Erklärungskraft anzeigen. In diesem Modell ist das R^2 der Markenwahrnehmung 0,585 und das des Markenverhaltens 0,578 (vgl. Tab. 35, Anhang S. A33). Das bedeutet, dass das Modell knapp 60 % der Varianz erklärt, was auf eine mittlere Erklärungskraft hindeutet (vgl. Chin et al., 1998, S. 325).

Tabelle 22: Ergebnisse des Messmodells

Quelle: Eigene Darstellung

Konstrukt	Faktorladungen	DEV	Cronbachs Alpha	Composite Reliability	Fornell-Larcker
Wahrnehmung von Greenwashing	0,721 – 0,875	64,4%	0,888	0,915	0,803
Rep. Produkte	0,774 – 0,832	63,8%	0,810	0,876	0,799 > 0,642
Rep. Innovation	0,587 – 0,849	57,7%	0,630	0,800	0,760 > 0,585
Rep. Arbeitsumfeld	0,896 – 0,942	83,6%	0,902	0,938	0,914 > 0,684
Rep. Leitung	0,861 – 0,901	77,2%	0,852	0,910	0,878 > 0,684
Rep. Ges. Einb.	0,834 – 0,857	72,0%	0,806	0,885	0,848 > 0,718
Rep. Führung	0,736 – 0,808	57,9%	0,757	0,846	0,761 > 0,615
Rep. Performance	0,795 – 0,851	67,4%	0,762	0,861	0,821 > 0,481
MW	0,824 – 0,884	72,2%	0,808	0,886	0,850 > 0,734
MV	0,846 – 0,939	81,9%	0,889	0,931	0,905 > 0,734

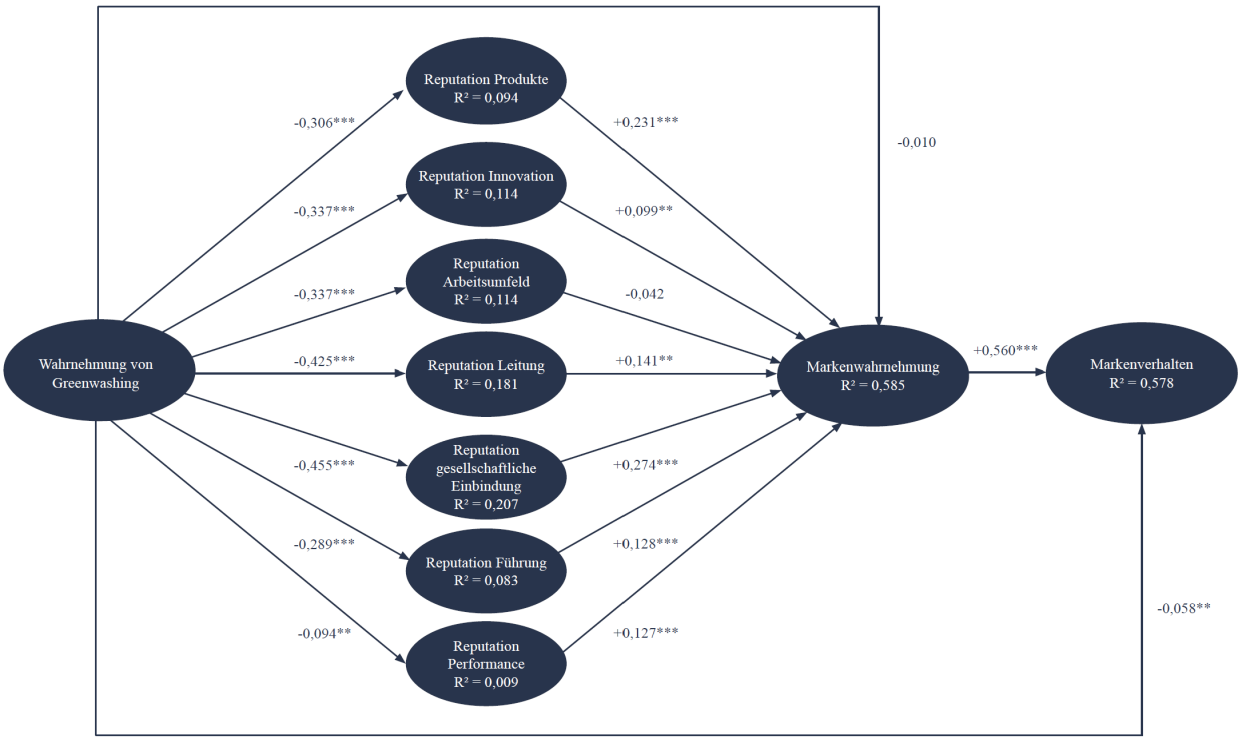


Abbildung 6: Strukturmodell mit Pfadkoeffizienten und Bestimmtheitsmaß R^2

*** p-Wert < 0,01
** p-Wert < 0,05
* p-Wert < 0,1
Quelle: Eigene Darstellung

Q^2 ist eine ähnliche Kennzahl wie R^2 und gibt aufgrund des Stone-Geisser-Tests Auskunft über die voraussagbare Relevanz der abhängigen Konstrukte. Sie sollte stets > 0 sein, um eine Vorhersagekraft zu haben (vgl. Chin, 2010, S. 680), was in diesem Modell der Fall ist (vgl. Tab. 35, Anhang S. A33). Da das Q^2 der Markenwahrnehmung 0,409 und das des Markenverhaltens 0,464 beträgt, besitzt dieses PLS-Pfadmodell eine große voraussagbare Relevanz (vgl. **?**, S. 208).

Zusammenfassend lässt sich sagen, dass die Auswertung des Messmodells und die strukturellen Zusammenhänge die vorgestellten kausalen Zusammenhänge zwischen der von den Probanden wahrgenommenen Stärke von Greenwashing, der Reputation und der daraus resultierenden Markenwahrnehmung sowie dem Markenverhalten belegen.

4.3. Überprüfung und Diskussion der Forschungshypothesen

In diesem Abschnitt sollen nun die im dritten Teil dieser Arbeit aufgestellten Hypothesen (vgl. Tab. 1) überprüft und die festgestellten Ergebnisse interpretiert werden. Dazu werden zunächst die Hypothesen H_1 und H_5 betrachtet. Aufgrund der Signaltheorie wird in H_1 angenommen, dass die unterschiedlichen CSR-Initiativen unterschiedliche Effekte auf die Reputation bzw. nachgelagert auch auf die Markenwahrnehmung und das Markenverhalten haben. Da frühere Studien (vgl. McQuarrie & Mick, 2003, S. 218; A. A. Mitchell, 1986, S. 16-21) gezeigt haben, dass visuelle Kommunikationsmethoden mehr Aufmerksamkeit erregen, reichere Schlussfolgerungen hervorrufen und überzeugender sind als schriftliche Aussagen wird in H_5 vermutet, dass Umweltlabel grundsätzlich eine stärkere Reaktion bei den Konsumenten hervorrufen als rein schriftliche CSR-Initiativen. Bei den in dem Experiment dieser Studie verwendeten CSR-Initiativen „Label“ und „Keine Vagheit“ können allerdings kaum Unterschiede beobachtet werden. Tendenziell wird die Reputation von den Probanden der Gruppe 3, also der CSR-Initiative mit dem Umweltlabel „Der blaue Engel“ besser bewertet. Auch die Faktoren Markenwahrnehmung (Gruppe 3: Ø 3,88; Gruppe 6: Ø 3,61) und Markenverhalten (Gruppe 3: Ø 3,81; Gruppe 6: Ø 3,54) erzielen hier etwas bessere Ergebnisse. Diese Unterschiede sind allerdings nicht signifikant, weshalb die beiden Hypothesen H_1 und H_5 abgelehnt werden (vgl. Tab. 5 & 6). Die Gründe dafür können vielschichtig sein. Zum einen ist es möglich, dass „Der blaue Engel“ als Label in der Werbeanzeige etwas untergegangen ist. Dadurch könnte der Fokus der Probanden mehr auf dem beworbenen Schuh an sich sowie dem dazugehörigen Text gelegen haben. Weiterhin könnte das Umweltlabel zwar erkannt worden sein, ihm wurde aber keine besondere Bedeutung zugeschrieben, da heute eine Vielzahl von Produkten mit irgendeiner Art von Umweltlabel beworben werden, wodurch die höhere Aufmerksamkeit egalisiert wird.

Ähnliche Ergebnisse lassen sich bei der Überprüfung der Hypothesen H_6 und H_9 erkennen. Auch hier ist zunächst angenommen worden, dass sich die Effekte der beiden Greenwashing-Methoden auf die Reputation von Adidas unterscheiden (H_6) und die Methode der Verwendung falscher Label stärkere Reaktionen nach sich zieht (H_9). Entgegen der Erwartungen können allerdings auch hier in der Post-hoc Analyse keine signifikanten Unterschiede festgestellt werden, wodurch beide Hypothesen abgelehnt werden. Verschiedene Greenwashing-Methoden führen in dieser Studie nicht zu unterschiedlichen Ergebnissen (vgl. Tab. 5 & 6). Auch hier besteht die Möglichkeit, dass das Fake-Label nicht allzu stark wahrgenommen wurde und der Fokus der Probanden eher auf den Sneakers sowie dem Text gelegen hat. Außerdem könnte die inflationäre Verwendung von Umweltsiegeln die Bedeutung ebendieser für die Konsumenten verringert haben.

Analog zu den bisher betrachteten Hypothesen werden nun auch noch die Hypothesen H_{10} und H_{12} überprüft. Hier ist angenommen worden, dass die Greenwashing-Skandale über die Verwendung falscher Label bzw. vager Behauptungen unterschiedliche Effekte auf die Reputation haben (H_{10}). Hypothese H_{12} folgt der Annahme, dass vage Behauptungen zwar irreführend sind, aber vermutlich nach einem Skandal von Konsumenten eher verziehen werden als skrupellos gefälschte Umweltzeichen. Wie sowohl bei den CSR-Initiativen als auch den Greenwashing-Methoden sind allerdings auch hier keine Unterschiede in Bezug auf die Reputation sowie die Markenwahrnehmung oder das Markenverhalten erkennbar (vgl. Tab. 5 & 6). Somit werden auch die Hypothesen H_{10} und H_{12} abgelehnt. Darüber hinaus wird auch die Reaktion auf einen Greenwashing-Skandal analysiert. Anhand der Tabelle 21 kann erkannt werden, dass der Faktor „Reaktion Skandal Vertrauen“ von den Probanden der Gruppe 4 (Ø 3,12) signifikant höher bewertet wurde als von denen der Gruppe 7 (Ø 2,73). Bei dem Faktor „Reaktion Skandal Neubewertung“ gibt es keine signifikanten Unterschiede. Der Greenwashing-Skandal über die Verwendung falscher Label bei Adidas hat also etwas stärker negative Reaktionen zur Folge als der Skandal über vage Behauptungen. Insgesamt sind die Effekte, vor allem auf die Reputation, allerdings sehr ähnlich.

Im nächsten Abschnitt werden die Hypothesen H_2 und H_7 untersucht. Dazu sind die Gruppen 2 & 5, 3 & 6 sowie 4 & 7 zusammengefasst worden. H_2 folgt dabei den Ergebnissen verschiedener Studien, dass CSR-Initiativen dazu beitragen können, die Unternehmensreputation (vgl. Saeidi et al., 2015, S. 347-348; Miles & Covin, 2000, S. 308-309) und dem nachgelagert die Markenwahrnehmung und das Markenverhalten (vgl. Brown & Dacin, 1997, S. 68) zu verbessern. Demgegenüber wird in H_7 behauptet, dass Greenwashing negativ auf die Gesamtleistungsindikatoren von Unternehmen wirkt. So lassen Studien wie die von Y.-S. Chen und Chang (2013, S. 497) vermuten, dass Greenwashing negative Effekte auf die Reputation zur Folge hat. Tatsächlich zeigen die Werte aus den Tabellen 7 und 8 keine signifikanten Unterschiede zwischen der Kontrollgruppe, den CSR-Initiativen und den Greenwashing-Methoden. Zwar lassen sich stets leicht höhere Mittelwerte in der Spalte der CSR-Initiativen im Vergleich zu der Kontrollgruppe und den Greenwashing-Methoden ablesen, diese Unterschiede sind allerdings nicht signifikant. Die Werte der Greenwashing-Methoden sind darüber hinaus nahezu identisch mit denen der Kontrollgruppe. Bei den Faktoren Markenwahrnehmung und Markenverhalten ergeben sich sogar exakt die gleichen Mittelwerte (vgl. Tab. 8). Daher werden die Hypothesen H_2 und H_7 abgelehnt. Folglich muss diskutiert werden, wieso sich weder bei den CSR-Initiativen positive noch bei den Greenwashing-Methoden negative Effekte beobachten lassen.

Als erster Grund dafür kann die hohe Skepsis der Probanden gegenüber den CSR-Initiativen aufgeführt werden. Mit 109 von 152 Probanden denken die meisten Probanden, das Nachhaltigkeitsprogramm der Marke Adidas sei stark strategisch motiviert (vgl. Tab. 13). Eine starke ökologische Motivation nehmen hingegen lediglich 49 der 152 Probanden wahr (vgl. Tab. 14). Viele Probanden vertrauen also nicht darauf, dass Adidas tatsächlich ein großes ökologisches/soziales Interesse hat.

Eine weitere Begründung ist, dass die Probanden kein allzu gutes Wissen in Bezug auf Umweltlabel vorweisen können. Bei den in der Umfrage gezeigten Symbolen konnte häufig nicht richtig zugeordnet werden, ob es sich um ein zertifiziertes oder ein gefälschtes Label handelt. Insgesamt haben lediglich 116 der insgesamt 517 Probanden alle vier Zeichen korrekt erkannt. Während die tatsächlichen Fake-Labels auch zumeist als solche identifiziert worden sind (Symbol 1 wurde von 83,4% korrekt erkannt; Symbol 4 wurde von 79,9% korrekt erkannt) zeigt sich bei den beiden zertifizierten Labels die Skepsis und Verwirrung der Probanden. So wurde Symbol 3 (Der blaue Engel) von 30,8% der Probanden als Fake wahrgenommen und Symbol 3 (EU-Bio-Logo) mit 52% der Probanden sogar von über der Hälfte (vgl. Tab. 23).

Außerdem können Fehleinschätzungen bei der Wahrnehmung von Greenwashing als Grund angeführt werden. Die Tabelle 9 zeigt die Mittelwerte dieser Wahrnehmung von den Probanden der CSR- und der Greenwashing-Gruppen. Es lässt sich zwar ein signifikanter Unterschied erkennen ($p < 0,05$), jedoch liegen die Werte relativ nah beieinander. Viele Probanden konnten demnach die CSR-Initiativen und Greenwashing-Methoden nicht eindeutig als solche erkennen. Insgesamt zeigen die Mittelwerte, welche sich zwischen 3,27 und 3,63 bewegen, wie hoch die Skepsis und Verwirrung in Bezug auf Green Marketing allgemein ist (vgl. Tab. 9). Dies unterstreichen auch die Tabellen 11 und 12. So haben 41 von 148 Probanden, denen eine der beiden CSR-Initiativen gezeigt wurde, Greenwashing dennoch stark wahrgenommen. Bei den Probanden aus den Greenwashing-Gruppen sind es mit 49 von 159 nur wenige mehr (vgl. Tab. 11 & 12).

Eine hohe Skepsis, ausbaufähiges Umweltwissen sowie ein daraus folgendes fehlerhaftes Erkennen von CSR und Greenwashing haben folglich dazu geführt, dass sich zwischen der Kontrollgruppe, den CSR-Initiativen und den Greenwashing-Methoden keine signifikanten Unterschiede erkennen lassen.

Nun werden die Greenwashing-Skandale in Form der Hypothese H_{11} überprüft. Hier wird vermutet, dass Greenwashing-Skandale die negativen Effekte von Greenwashing auf die Reputation von Adidas verstärken. Die Greenwashing-Methoden haben, wie soeben besprochen, keine signifikanten negativen Effekte auf die Reputation. Bei den Greenwashing-Skandalen zeigen sich hingegen deutlich niedrigere Mittelwerte in Bezug auf die RepTrak-Faktoren, die Markenwahrnehmung und auch das Markenverhalten. Insbesondere die RepTrak-Faktoren Leitung (Ø 2,58) und gesellschaftliche Einbindung (Ø 2,57) werden erheblich schlechter bewertet als in den anderen Gruppen (vgl. Tab. 7). Die Hypothese H_{11} kann also bestätigt werden. Die Zeitungsartikel in der Umfrage haben dafür gesorgt, dass die Probanden keine Unsicherheit mehr darüber haben, ob es sich bei der Werbeanzeige um CSR oder Greenwashing handelt. Dadurch haben viele Probanden das dargestellte Fehlverhalten von Adidas verurteilt und die Reputation deutlich schlechter bewertet. Die Markenwahrnehmung und das Markenverhalten sind zwar auch deutlich schlechter bewertet worden, allerdings sind die Mittelwerte beider Faktoren noch immer über dem Wert 3 (MW: Ø 3,30; MV: Ø 3,09). Es scheint, als würde die gute Reputation von Adidas die Effekte eines Greenwashing-Skandals auf das tatsächliche Verhalten der Konsumenten abdämpfen. Trotz des offensichtlichen ökologischen Fehlverhaltens besteht so bspw. noch immer eine akzeptable Kaufbereitschaft (vgl. Tab. 8).

Im nächsten Schritt werden die Auswirkungen der Wahrnehmung von Greenwashing untersucht. In Hypothese H_3 wird behauptet, dass CSR-Initiativen negative Effekte auf die Unternehmensreputation haben, wenn sie fälschlicherweise für Greenwashing gehalten werden. Die Tabelle 11 zeigt die Unterschiede der Mittelwerte zwischen den Probanden aus den CSR-Gruppen, die Greenwashing gering wahrgenommen haben und denen, die es stark wahrgenommen haben. Darüber hinaus werden noch die Mittelwerte der Kontrollgruppe angegeben. So wird ersichtlich, dass die CSR-Initiativen zu durchweg höheren Mittelwerten geführt haben, sofern sie nicht fälschlicherweise als Greenwashing-Methoden wahrgenommen werden. Insbesondere die RepTrak-Faktoren Arbeitsumfeld (Ø 3,11) und gesellschaftliche Einbindung (Ø 3,35) sind deutlich besser bewertet worden als in der Kontrollgruppe (Ø 2,76; Ø 2,97).

Auf der anderen Seite sinken sowohl die Werte der RepTrak-Faktoren als auch der Markenwahrnehmung und dem Markenverhalten, wenn Greenwashing stark wahrgenommen wird. Vor allem bei dem Faktor Rep. Leitung (Ø 2,62) und erneut dem Faktor Rep. gesellschaftliche Einbindung (Ø 2,62) lassen sich erheblich schlechtere Werte als bei der Kontrollgruppe (Ø 2,94; Ø 2,97) erkennen (vgl. Tab. 11). Somit kann die Hypothese H_3 bestätigt werden. CSR-Initiativen, die fälschlicherweise als Greenwashing wahrgenommen werden, haben negative Effekte auf die Reputation von Adidas.

Mit der Hypothese H_8 wird genau die gegensätzliche Wirkung überprüft. Diese behauptet, Greenwashing-Methoden haben positive Effekte auf die Unternehmensreputation, wenn sie fälschlicherweise für CSR-Initiativen gehalten werden. Zur Überprüfung wird davon ausgegangen, dass die Probanden eine Greenwashing-Methode für eine CSR-Initiative halten, wenn sie Greenwashing gar nicht oder nur gering wahrnehmen. Die Ergebnisse sind hier ähnlich wie die der Hypothese H_3. Wird Greenwashing korrekterweise stark wahrgenommen, verringern sich die Mittelwerte aller Faktoren, vor allem die Rep. Leitung sowie Rep. gesellschaftliche Einbindung. Wird die Greenwashing-Methode aber für CSR gehalten, können beinahe für jeden Faktor höhere Werte als in der Kontrollgruppe verzeichnet werden (vgl. Tab. 12). Aus diesem Grund kann auch die Hypothese H_8 bestätigt werden. Werden Greenwashing-Methoden fälschlicherweise für CSR gehalten zeigen sich positive Effekte auf die Reputation von Adidas.

Um darüber hinaus auch die Kausalbeziehungen zwischen der Wahrnehmung von Greenwashing, den RepTrak-Dimensionen sowie der Markenwahrnehmung und dem Markenverhalten analysieren zu können, wurde im vorigen Kapitel noch eine strukturelle Gleichungsmodellierung durchgeführt.

Tabelle 23: Ergebnisse der Überprüfung der EK

Quelle: Eigene Darstellung

Symbol	Als zertifiziert wahrgenommen		Als Fake wahrgenommen	
	n	%	n	%
Symbol 1 (Fake)	86	16,6	431	83,4
Symbol 2 (Zertifiziert)	358	69,2	159	30,8
Symbol 3 (Zertifiziert)	248	48,0	269	52,0
Symbol 4 (Fake)	104	20,1	413	79,9

Hier lassen sich zunächst einmal relativ starke, signifikante Beziehungen zwischen der Wahrnehmung von Greenwashing und den RepTrak-Faktoren erkennen. Insbesondere auf die Leitung (- 0,425) sowie die gesellschaftliche Einbindung (- 0,455) bestehen starke negative Einflüsse (vgl. Abb. 6). In diesen Faktoren werden ethische, gemeinnützige Aspekte betrachtet. Greenwashing wird also als unethisch und unfair erkannt. Die Rep. Performance wird hingegen nicht signifikant beeinflusst, die wahrgenommene finanzielle Stärke von Adidas scheint also aufgrund von Greenwashing nicht schlechter wahrgenommen zu werden. Weiterhin zeigen sich nur sehr schwache negative Effekte auf die Markenwahrnehmung und das Markenverhalten. Diese Beziehungen sind zudem nicht signifikant. Der Reputation nachgelagerte Variablen, wie das Markenvertrauen, die Kaufabsicht oder die Preisakzeptanz werden durch wahrgenommenes Greenwashing demnach zumindest kurzfristig nicht signifikant negativ beeinflusst. Langfristig können möglicherweise auch hier stärkere negative Effekte beobachtet werden, insbesondere, wenn die Konsumenten weiterhin Fake-Label, irreführende Aussagen oder sonstige Greenwashing-Methoden bei Adidas wahrnehmen. In jedem Fall zeigt sich, dass eine gute Reputation, wie die von Adidas, den Schaden auf die Markenwahrnehmung und das Markenverhalten abschwächen kann. Insbesondere die RepTrak-Faktoren Produkte und gesellschaftliche Einbindung haben dabei verhältnismäßig starke ($> 0,2$) Einflüsse auf die Markenwahrnehmung. Da die Markenwahrnehmung hier auch als einflussreiche Mediatorvariable (+ 0,560) zwischen den RepTrak-Faktoren und dem Markenverhalten agiert, gelten diese Einflüsse indirekt auch für das Markenverhalten (vgl. Abb. 6). Sowohl die Produkte einer Marke als auch die gesellschaftliche Einbindung sollten also besonders positiv wahrgenommen werden, um die Markenwahrnehmung und das Markenverhalten zu stärken und negative Effekte aufgrund von wahrgenommenem Greenwashing abzuschwächen.

Darüber hinaus wird die Hypothese H_4 überprüft. Dazu werden die wahrgenommenen Motivationen für die dargestellten CSR-Initiativen von Adidas untersucht und die Auswirkungen analysiert. Aufgrund der Forschungsergebnisse vorangegangener Untersuchungen wird davon ausgegangen, dass eine wahrgenommene ökologische Motivation stärkere positive Auswirkungen auf die Effekte von CSR-Initiativen hat als eine strategische Motivation. Tabelle 13 zeigt die unterschiedlichen Mittelwerte der RepTrak-Faktoren, je nachdem ob von den Probanden eine geringe oder eine starke strategische Motivation wahrgenommen wurde. Tendenziell lassen sich hier bessere Werte erkennen, wenn die strategische Motivation stark wahrgenommen wurde, signifikant besser ist aber lediglich der Faktor Rep. Führung bewertet worden.

Demgegenüber zeigt Tabelle 14 die unterschiedlichen Mittelwerte bei gering bzw. stark wahrgenommener ökologischer Motivation. Wird eine starke ökologische Motivation erkannt, werden vor allem die RepTrak-Faktoren Arbeitsumfeld (+ 0,69), Leitung (+ 1,02) und gesellschaftliche Einbindung (+ 0,97) deutlich besser bewertet. Die Hypothese H_4 kann bestätigt werden. Eine erkennbare ökologische Motivation ist entscheidend für den Erfolg von CSR-Initiativen. Marken sind dabei in aller Regel nicht ausschließlich strategisch oder ökologisch motiviert. Es bedarf einer Kombination aus beidem, um CSR langfristig und sinnvoll zu etablieren. Laut Chandler (2016, S. 248) ist CSR nur dann effizient, wenn es auch in die strategische Planung eines Unternehmens eingegliedert wird. Den meisten Konsumenten ist bewusst, dass Unternehmen wirtschaftlich agieren müssen, um im internationalen Wettbewerb überleben zu können. Diese wirtschaftliche Motivation wird auch nicht verurteilt, vielmehr stärkt sie sogar die Reputation in einigen Faktoren. Wichtig ist, dass neben der strategischen auch die ökologische Motivation klar erkennbar ist. Die Konsumenten müssen dem Unternehmen glauben, dass es der Gesellschaft tatsächlich etwas zurückgeben will und der Fokus nicht allein auf den möglichen finanziellen Vorteilen liegt. Wird dies nicht wahrgenommen, können keine signifikanten Effekte von CSR-Initiativen erwartet werden (vgl. Tab. 14).

Weiterhin behauptet Hypothese H_{13}, dass ein hohes Umweltinvolvement dazu führt, dass CSR-Initiativen und Greenwashing-Methoden eher als solche erkannt werden. Dazu wird sowohl das Umweltbewusstsein (EC) als auch das Umweltwissen (EK) betrachtet. In Bezug auf das Umweltbewusstsein können dabei keine signifikanten Unterschiede erkannt werden. Die Tabellen 15 und 17 zeigen, dass die Probanden mit einem hohen Umweltbewusstsein weder CSR-Initiativen noch Greenwashing-Methoden eher als solche erkennen konnten, als die Probanden mit einem geringeren

Umweltbewusstsein. Bei der Untersuchung des Umweltwissens können hingegen durchaus Unterschiede festgestellt werden. So wurde Greenwashing bei den CSR-Initiativen signifikant schwächer und bei den Greenwashing-Methoden signifikant stärker wahrgenommen, wenn die Probanden über ein hohes Umweltwissen verfügen (vgl. Tab. 16 & 18). Diese Ergebnisse passen zu der Überlegung, dass EC als eher subjektiver Indikator das Bewusstsein der Konsumenten für Umweltbelange angibt, während EK die Rolle als objektiver Indikator für ihre Fähigkeit, Greenwashing in der Werbung wahrzunehmen, einnimmt (vgl. Schmuck, Matthes & Naderer, 2018, S. 131). Darüber hinaus könnte die persönliche Selbsteinschätzung der Probanden nach ihrem Umweltbewusstsein zu Verzerrungen geführt haben, weshalb sich keine signifikanten Gruppenunterschiede erkennen lassen. Zusammengefasst kann die Hypothese H_{13} dennoch bestätigt werden.

Laut Hypothese H_{14} führt weiterhin ein junges Alter dazu, dass CSR-Initiativen und Greenwashing-Methoden eher als solche erkannt werden. Verschiedene Studien (vgl. Besel et al., 2017, S. 67-70; Chan & Leung, 2006, S. 436) haben ergeben, dass junge Menschen, insbesondere Universitätsstudenten, sehr informiert über Fragen im Zusammenhang mit Umwelt und Umweltkommunikation sind. Während bei den CSR-Initiativen keine signifikanten Unterschiede festgestellt werden können (vgl. Tab. 19), sind die Greenwashing-Methoden jedoch deutlich häufiger als solche erkannt worden. Bei der Wahrnehmung von Greenwashing zeigte sich hier ein Mittelwert von 3,71 bei Probanden bis 25 Jahren und ein Mittelwert von 3,33 bei den Probanden ab 26 Jahren (vgl. Tab. 20). Die Hypothese H_{14} kann demzufolge zumindest teilweise bestätigt werden.

Außerdem wurde in einer Studie von Torelli et al. (2020, S. 416-417) beobachtet, dass auch das Geschlecht der Konsumenten ein wichtiger Einflussfaktor sein kann. Weibliche Probanden haben Greenwashing demnach häufiger erkannt als männliche. Ob dies auch in der vorliegenden Studie der Fall ist, wird anhand der Hypothese H_{15} überprüft. Zwar wurde die Wahrnehmung von Greenwashing von den Frauen der Greenwashing-Gruppen mit einem Mittelwert von 3,63 etwas höher bewertet als von den Männern (Ø 3,49), dieser Unterschied ist allerdings nicht auf das 5%-Niveau signifikant (vgl. Tab. 32, Anhang S. A30). Die Hypothese H_{15} wird demnach abgelehnt.

Abschließend wird noch Hypothese H_{16} überprüft. Hier wird behauptet, Greenwashing-Skandale haben bei Frauen stärkere negative Effekte als bei Männern. Demnach wird hier die Meinung über das Unternehmen bei Frauen langfristiger geschädigt, wodurch sie das Unternehmen in Zukunft eher meiden als Männer (vgl. Torelli et al., 2020, S. 416-417). Wie in der Tabelle 33 (vgl. Anhang S. A31) zu sehen, hat sich diese These hier nicht bestätigt. Bei beiden Faktoren der Reaktion auf einen Greenwashing-Skandal lassen sich keine signifikanten Unterschiede erkennen. Hypothese H_{16} wird abgelehnt. Insgesamt ist erkennbar, dass die Reaktionen auf die Skandale verhältnismäßig mild ausfallen. Obwohl in den Zeitungsartikeln erhebliche Missstände bei der Produktion aufgedeckt wurden, liegt überhaupt nur ein Mittelwert über dem Wert 3 (vgl. Tabelle 21). Eine mögliche Begründung ist, dass Adidas als Marke sehr geschätzt wird. Diese hohe Reputation schwächt demnach (zumindest kurzfristig) größere negative Effekte auf das Verhalten der Konsumenten ab.

Die Tabelle 24 zeigt die Ergebnisse aus den Hypothesenprüfungen.

5. Implikationen und Limitationen

5.1. Implikationen für die Marketingpraxis

Die Ergebnisse dieser Analyse können für Unternehmen nützlich sein, um die Konsequenzen von unterschiedlichen „grünen“ Kommunikationsstrategien besser verstehen zu können. Anhand der dadurch möglichen optimierten Planung und Umsetzung der Kommunikation können so Misstrauen und Legitimationsverluste eher vermieden werden.

Zunächst einmal zeigt diese Studie, dass es weniger darauf ankommt, welche CSR-Initiative oder Greenwashing-Methode verwendet wird. Es können keine signifikanten Unterschiede zwischen den jeweils verwendeten Methoden festgestellt werden. Überraschenderweise können weiterhin auch zwischen der Kontrollgruppe, den beiden CSR-Gruppen und den beiden Greenwashing-Gruppen keine Unterschiede in der Grundgesamtheit beobachtet werden. Signifikant negative Effekte zeigen sich lediglich bei den Greenwashing-Skandalen. Als Gründe hierfür sind insbesondere die hohe Skepsis der Konsumenten bezüglich der Motivation, die starke Wahrnehmung von Greenwashing in allen Stichproben sowie ein durchwachsenes Umweltwissen identifiziert worden. Diese Gründe beeinflussen sich dabei offensichtlich gegenseitig. Bspw. führt fehlendes Wissen über Umweltlabel zu einer höheren Skepsis und Unsicherheiten bei der Erkennung von Greenwashing.

Die Skepsis bezüglich der Motivation zeigt sich unter anderem dadurch, dass die strategische Motivation deutlich häufiger wahrgenommen wurde als die ökologische. So wurde bei den CSR-Initiativen eine starke strategische Motivation von 74%, eine starke ökologische Motivation allerdings lediglich von 33% der Probanden wahrgenommen (vgl. Tab 13 & 14). Während eine starke strategische Motivation zwar akzeptiert wird, aber keine positiven Effekte verursacht, kann die ökologische Motivation als wesentlicher Treiber für den Erfolg von CSR-Initiativen erkannt werden. Mit Ausnahme des RepTrak-Faktors Performance werden alle anderen Faktoren signifikant besser bewertet, wenn die CSR-Initiative als stark ökologisch motiviert wahrgenommen wurde. Diese ökologische Motivation muss beim Green Marketing daher unbedingt erkennbar sein und glaubhaft gemacht werden. Die strategische Motivation ist ebenso notwendig, um CSR erfolgreich zu etablieren und daraus auch Wettbewerbsvorteile ziehen zu können (vgl. Chandler, 2016, S. 248). Für die Konsumenten ist es aber entscheidend, dass die finanziellen Beweggründe nicht an erster Stelle stehen und ein tatsächlich nachhaltiger bzw. umweltfreundlicher Grundgedanke verfolgt wird.

Tabelle 24: Übersicht über die bestätigten bzw. abgelehnten Hypothesen

Quelle: Eigene Darstellung

Hypothese		Ergebnis
H_1	Unterschiedliche CSR-Initiativen haben unterschiedliche Effekte auf die Unternehmensreputation.	Abgelehnt
H_2	CSR-Initiativen haben insgesamt positive Effekte auf die Unternehmensreputation.	Abgelehnt
H_3	CSR-Initiativen haben negative Effekte auf die Unternehmensreputation, wenn sie fälschlicherweise für Greenwashing gehalten werden.	Bestätigt
H_4	Als intrinsisch motiviert wahrgenommene CSR-Initiativen haben stärkere positive Effekte auf die Unternehmensreputation als solche, die als extrinsisch motiviert wahrgenommen werden.	Bestätigt
H_5	CSR-Initiativen mit einem Umweltlabel haben stärkere positive Effekte auf die Unternehmensreputation als CSR-Initiativen ohne Umweltlabel.	Abgelehnt
H_6	Unterschiedliche Greenwashing-Methoden haben unterschiedliche Effekte auf die Unternehmensreputation.	Abgelehnt
H_7	Greenwashing-Methoden haben insgesamt negative Effekte auf die Unternehmensreputation.	Abgelehnt
H_8	Greenwashing-Methoden haben positive Effekte auf die Unternehmensreputation, wenn sie fälschlicherweise für CSR-Initiativen gehalten werden.	Bestätigt
H_9	Greenwashing-Methoden mit einem Fake-Label haben stärkere Effekte auf die Unternehmensreputation als solche mit vagen Formulierungen.	Abgelehnt
H_{10}	Unterschiedliche Greenwashing-Skandale haben unterschiedliche Effekte auf die Unternehmensreputation.	Abgelehnt
H_{11}	Greenwashing-Skandale verstärken die negativen Effekte von Greenwashing auf die Unternehmensreputation.	Bestätigt
H_{12}	Greenwashing-Skandale zu einem Fake-Label haben stärkere negative Effekte auf die Unternehmensreputation als solche zu vagen Formulierungen.	Abgelehnt
H_{13}	Ein hohes Umweltinvolvement führt dazu, dass CSR-Initiativen und Greenwashing-Methoden eher als solche erkannt werden.	Bestätigt
H_{14}	Ein junges Alter führt dazu, dass CSR-Initiativen und Greenwashing-Methoden eher als solche erkannt werden.	Teilweise bestätigt
H_{15}	Frauen können CSR-Initiativen und Greenwashing-Methoden häufiger als solche identifizieren als Männer.	Abgelehnt
H_{16}	Greenwashing-Skandale haben bei Frauen stärkere negative Effekte als bei Männern.	Abgelehnt

Die starke Wahrnehmung von Greenwashing zeigt sich in den Mittelwerten, welche in den jeweils zwei CSR- sowie Greenwashing-Gruppen zwischen 3,27 und 3,63 liegen, wobei lediglich ein geringer signifikanter Unterschied zwischen den CSR-Initiativen und den Greenwashing-Methoden festgestellt werden kann (vgl. Tab. 9). Demnach wurde Greenwashing teilweise bei CSR-Initiativen wahrgenommen, während Greenwashing-Methoden als CSR beurteilt wurden (vgl. Tab 11 & 12). Dabei fällt auf, dass CSR-Initiativen, welche als Greenwashing-Methode wahrgenommen wurden, negative Effekte auf die Reputation haben und Greenwashing-Methoden, welche als CSR-Initiative wahrgenommen wurden, positive. Insbesondere die RepTrak-Faktoren Leitung und gesellschaftliche Einbindung hängen dabei sehr von der wahrgenommenen Stärke von Greenwashing ab (vgl. Abb. 6).

Die Wahrnehmung von Greenwashing sowie die Wahrnehmung der ökologischen Motivation können folglich als wesentliche Treiber für die Effekte von Green Marketing auf die Reputation beobachtet werden. Unternehmen müssen daher sicherstellen, dass nachhaltige Unternehmenspraktiken sehr transparent sind und als wesentliches Ziel in die Mission und Vision eingebaut werden. Wenn die intrinsische Motivation zu ökologischem Handeln tief in der Marke verwurzelt ist, wird sich auch die Wahrnehmung der Motivation durch die Konsumenten positiv verändern und grüne Werbekampagnen dieser Marke werden vermutlich deutlich häufiger für CSR und deutlich seltener für Greenwashing gehalten.

Insgesamt scheint es unumgänglich, zu versuchen, die Skepsis und Verwirrung der Konsumenten gegenüber Green Marketing zu mildern. Solange dies nicht geschieht, werden tatsächliche CSR-Initiativen nie ihr vollständiges Potential erreichen können, was nicht nur schlecht für die engagierten Unternehmen ist, sondern auch für die Gesellschaft im Allgemeinen. Je lohnenswerter der ökonomische Part des CSR ist, desto mehr Unternehmen werden sich daran beteiligen

und es sich auch selbst leisten können, der Gesellschaft etwas zurückzugeben. Sollte sich die Skepsis allerdings nicht mindern oder sogar noch erhöhen, weil sich Greenwashing weiter verbreitet, werden CSR-Initiativen womöglich schon bald nicht mehr tragbar sein, da sie zu häufig erst gar nicht beachtet oder für Greenwashing gehalten werden. Eine solche Entwicklung könnte den gesamten Markt für nachhaltige Produkte gefährden und gilt es unbedingt zu vermeiden.

Aufgrund dieser Ergebnisse stellt sich die Frage, was nachhaltige Unternehmen dafür tun können die Skepsis der Konsumenten zu verringern und dafür zu sorgen, dass ihre Produkte ohne Zweifel von solchen unterschieden werden können, die nur vorgeben, nachhaltig zu sein. Das Problem und eine mögliche Lösung liegen hier sehr nah beieinander. Aktuell führt die Masse an undurchsichtigen Umweltlabels dafür, dass viele Konsumenten nicht wissen, welches Label welche Aussagekraft hat, welche Kriterien dafür erfüllt sein müssen und welches überhaupt von einer unabhängigen Organisation zertifiziert wurde. Auch in dieser Studie ist häufig nicht richtig erkannt worden, welches Umweltlabel zertifiziert ist und welches lediglich erfunden wurde (vgl. Tab. 23).

Würde es aber nur sehr wenige, transparente und zertifizierte Label geben und würden alle anderen verboten und ihre Verwendung unter hohe Strafen gestellt werden, könnten sich womöglich viele Probleme lösen lassen. Unternehmen können sich dafür bei der Regierung für eine verbesserte Aufsicht und Regulierung einsetzen, um Greenwashing durch Wettbewerber zu vermindern (vgl. Kirchhoff, 2000, S. 404). Im Optimalfall erarbeiten die EU, Unternehmen und Umweltexperten gemeinsam wenige, transparente Umweltlabels. Die Aussagekraft und die zu erfüllenden Kriterien werden klar mit den Konsumenten kommuniziert, alle anderen Label werden untersagt und die Unternehmen werden regelmäßig von unabhängigen Organisationen auf die Einhaltung der Kriterien überprüft. Auf diese Weise könnte sich die Skepsis der Konsumenten signifikant verringern und umweltfreundliche Unternehmen, Produkte und Dienstleistungen könnten auf einen Blick als solche erkannt werden. Greenwashing würde dadurch sicherlich nicht sofort zu 100% besiegt werden, da Unternehmen noch immer irreführende Aussagen oder Naturbilder (Executional Greenwashing) in Verbindung mit ihrer Marke oder ihren Produkten präsentieren könnten. Anhand der zertifizierten Labels könnte aber verifiziert werden, ob Nachhaltigkeitsaspekte hier tatsächlich ausreichend beachtet wurden, oder eben nicht.

Bis es soweit ist, sollte Greenwashing zweifellos vermieden werden, sowohl aus ethischen als auch aus ökonomischen Gründen, da ein möglicher Skandal zu erheblichen Reputationsverlusten führen kann. Insbesondere die Reputations-Faktoren Leitung (Ø 2,58) und Einbindung in die Gesellschaft (Ø 2,57) sind in dieser Studie erheblich schlechter bewertet worden als in den anderen Gruppen. Die Markenwahrnehmung und das Markenverhalten sind zwar auch deutlich schlechter bewertet worden, allerdings zeigen sich hier noch immer solide Mittelwerte (vgl. Tab. 8). Die sehr gute Reputation von Adidas hat demnach offenbar die Effekte eines Greenwashing-Skandals auf das Verhalten der Konsumenten abdämpfen können. Obwohl in den Zeitungsartikeln offensichtliche ökologische Fehlverhalten dargestellt werden, sind so noch immer viele Konsumenten dazu bereit, Produkte der Marke Adidas zu erwerben. Diese Pufferfunktion der Reputation kann auch anhand des Strukturmodells verifiziert werden. Während sich starke negative Effekte von der Wahrnehmung von Greenwashing auf fast alle RepTrak-Faktoren zeigen, sind die Effekte auf die Markenwahrnehmung und das Markenverhalten lediglich minimal (vgl. Abb. 6).

Auch deswegen ist es für Unternehmen so wichtig, eine starke Reputation aufzubauen. Sie beeinflusst nicht nur kommerzielle Größen wie Umsatz oder Gewinn (vgl. Keh & Xie, 2009, S. 732), sondern kann auch vor negativen Effekten von wahrgenommenem Greenwashing und Skandalen schützen. Insbesondere bei den RepTrak-Faktoren Produkte und gesellschaftliche Einbindung zeigen sich dabei starke Einflüsse auf die Markenwahrnehmung und das Markenverhalten (vgl. Abb. 6).

Insgesamt sind die Ergebnisse hilfreich, um Greenwashing und die dabei verwendeten Aktionen und Strategien zu verstehen. Verantwortliche können ein umfassenderes Verständnis für die verschiedenen Motive, Treiber und Arten von Greenwashing erlangen und sind besser in der Lage, die Haupteffekte unterschiedlicher Umweltkommunikationen auf die Reputation zu berücksichtigen. Zwar lassen sich hier keine signifikanten Unterschiede zwischen der Kontrollgruppe, CSR und Greenwashing erkennen, aber selbstverständlich sollte dennoch schon aus ethischen Gründen jede Form von Greenwashing vermieden werden. Darüber hinaus ist es aber auch aus rein ökonomischer Perspektive wohl nur in den seltensten Fällen, und dann wahrscheinlich auch nur zeitweise, vorteilhaft für ein Unternehmen. Bei einem Skandal können schwerwiegende Konsequenzen wie Reputations- und Vertrauensverluste, Geldstrafen und ein Verlust von Marktanteilen auftreten. Auch wenn sich hier kurzfristig keine signifikanten negativen Effekte auf die Unternehmensreputation gezeigt haben solange kein Skandal aufkommt, so sorgt Greenwashing dennoch für Verwirrung und Skepsis bei den Konsumenten. Dadurch sinkt das Vertrauen in Green Marketing und der gesamte Markt für nachhaltige Produkte wird negativ beeinflusst, da tatsächlich nachhaltige Produkte nicht mehr eindeutig unterschieden werden können. Als potenzielle Lösung können wenige, transparente und regelmäßig kontrollierte Zertifikationen genannt werden.

5.2. Implikationen für die Marketingforschung

Die Auswertung dieser Studie hat einige Erkenntnisse über die Effekte von unterschiedlichen CSR-Initiativen, Greenwashing-Methoden sowie Greenwashing-Skandalen geliefert. Dennoch bleiben verschiedene Aspekte unbeantwortet. Im folgenden Abschnitt sollen die Fragestellungen diskutiert werden, die in der zukünftigen Marketingforschung untersucht werden könnten.

Zunächst einmal kann die wahrgenommene ökologische/soziale Motivation als wesentlicher Treiber für positive

Effekte des Green Marketing auf die Unternehmensreputation ausgemacht werden. Haben die Probanden in dieser Studie eine intrinsische Motivation für unternehmerische Nachhaltigkeit wahrgenommen, so ist die Reputation von Adidas signifikant besser bewertet worden (vgl. Tab. 14). Eine zentrale Frage ist, was die Konsumenten glauben lässt, dass eine Marke ökologisch/sozial motiviert handelt. Hier können keine signifikanten Unterschiede zwischen den einzelnen Stichproben festgestellt werden (vgl. Anhang Tab. 30). Weitere Studien könnten saher untersuchen, was genau Unternehmen tun können, um ihre intrinsische Motivation glaubhaft zu kommunizieren.

Weiterhin ist auch die Wahrnehmung von Greenwashing von zentraler Bedeutung. CSR-Initiativen haben im Vergleich mit der Kontrollgruppe nicht zu signifikant positiven Effekten geführt und Greenwashing-Methoden nicht zu signifikant negativen. Ein wesentlicher Grund dafür ist, dass die Probanden grundsätzlich skeptisch reagiert und demnach in allen Gruppen Greenwashing relativ stark wahrgenommen haben (vgl. Tab 9). Dabei wurden CSR-Initiativen häufig als Greenwashing wahrgenommen, während Greenwashing teilweise für CSR gehalten worden ist (vgl. Tab 11 & 12). Es sollte daher weiter detailliert untersucht werden, wie sich die Skepsis und Verwirrung der Konsumenten generell verringern lassen kann und wie genau einzelne Unternehmen überzeugend kommunizieren können, dass es sich bei ihrem Green Marketing um tatsächliches CSR und nicht um Greenwashing handelt. Als potenzieller Lösungsansatz ist die Einführung weniger, von Politik, Wirtschaft und unabhängigen Experten erarbeiteter Label genannt worden. Wie erfolgversprechend diese Lösung ist, gilt es aber anhand weiterer Studien zu erforschen. Womöglich würde auch eine Aufklärungskampagne helfen, um über die wichtigsten Label aufzuklären und Greenwashing als ernsthafte Problematik in den Köpfen der Konsumenten zu verankern. Dadurch könnten zertifizierte Label sowie Fake-Label vermutlich eher identifiziert werden.

Darüber hinaus werden in dieser Untersuchung lediglich die kurzfristigen Effekte der unterschiedlichen Werbeanzeigen und Skandale betrachtet. Um aussagekräftigere Ergebnisse liefern zu können, sollten auch die Langzeiteffekte betrachtet werden. So könnte untersucht werden, welche Auswirkungen weitere Skandale oder eine langfristige Wahrnehmung von Greenwashing hätten. Womöglich wird die Reputation ihre Pufferfunktion nicht unbefristet erfüllen können und auch das Verhalten der Konsumenten wird sich stärker negativ verändern. Andererseits wäre es spannend zu wissen, ob und wie schnell sich die Reputation eines Unternehmens wieder erholen kann, wenn die Konsumenten kein Greenwashing mehr wahrnehmen.

Zukünftige Forschung sollte außerdem die verschiedenen Methoden von Greenwashing verfeinern und erweitern. Hier sind lediglich die Verwendung falscher Label sowie die Vagheit betrachtet worden, es gibt allerdings noch viele weitere Methoden und auch CSR-Initiativen, deren Effekte untersucht werden können. Darunter zählt auch das Executional Greenwashing, bei dem naturanregende Elemente in der Werbung verwendet werden, um eine falsche Wahrnehmung der Umweltfreundlichkeit eines Unternehmens hervorzurufen (vgl. Parguel et al., 2015, S. 2).

Diese könnten dabei auch mit anderen Medien als einer Werbeanzeige kombiniert werden. So könnten bspw. auch TV-Spots oder Social Media Posts nachgestellt werden, um die Wirkungen auf die Reputation zu messen.

Es wäre auch sinnvoll, das durchgeführte Experiment in anderen Ländern und an einer Stichprobe durchzuführen, die hinsichtlich Alter und Bildungsniveau vielfältiger ist. Die Probanden in dieser Studie sind zum Großteil Studenten und junge Angestellte. Eine Stichprobe, die die Grundgesamtheit besser widerspiegelt, könnte wohl aufschlussreichere Ergebnisse liefern. Durch Untersuchungen in anderen Ländern könnten zudem gesellschaftliche und kulturelle Unterschiede herausgefunden werden, welche insbesondere für international agierende Unternehmen von großer Bedeutung sein können.

Weiterhin sollten auch andere Marken betrachtet werden. Adidas genießt insbesondere in Deutschland sehr hohe Beliebtheit, was sich an der betrachteten Einstellung gegenüber der Marke bemerkbar macht (vgl. Tab. 29, Anhang S. A29). Auch die relativ geringen Effekte auf die Markenwahrnehmung und das Markenverhalten (vgl. Tab. 8) sowie die milden Reaktionen auf einen Greenwashing-Skandal (vgl. Tab. 21) hängen vermutlich mit der starken Reputation von Adidas zusammen. Studien mit weniger angesehenen Marken könnten daher spannende Ergebnisse liefern. Insbesondere sollte das Experiment dieser Studie in Kombination mit einem Unternehmen aus einer umweltsensiblen Branche (z.B. Tabakproduktion oder Papierherstellung) durchgeführt werden. Laut Cho et al. (2009, S. 147-148) kann das Geschäftsfeld signifikante Einflüsse auf die Effekte von CSR-Initiativen und Greenwashing-Methoden haben. Auch die Reaktionen auf einen Greenwashing-Skandal könnten deutlich stärker ausfallen, wenn das Unternehmen in einer umweltsensiblen Branche tätig ist (vgl. Torelli et al., 2020, S. 416).

In dieser Studie wurden außerdem lediglich die Konsumenten als Stakeholder betrachtet. Der Forschung mangelt es aber bspw. noch an einem Verständnis der Auswirkungen von Greenwashing auf der individuellen Ebene innerhalb eines Unternehmens. Studien der Psychologie und des Organisationsverhaltens haben bereits Faktoren wie eine verbesserte psychische Gesundheit der Mitarbeiter, eine höhere Arbeitszufriedenheit und stärkeres organisatorisches Engagement in solchen Unternehmen festgestellt, die CSR betreiben (vgl. Robertson & Barling, 2013, S. 186-188). Zukünftige Forschungen sollten untersuchen, ob Unternehmen, die Greenwashing betreiben, einen weniger positiven oder sogar negativen Einfluss auf die Umweltbegeisterung und -aktionen der Mitarbeiter sowie auf die psychische Gesundheit, das Engagement und die Produktivität haben. Dies alles sind Faktoren, die sich negativ auf die Unternehmensleistung auswirken könnten. Darüber hinaus wäre es interessant zu untersuchen, wie Greenwashing die Handlungen von weiteren Stakeholdern wie politischen Entscheidungsträgern, Verbrauchergruppen, Lieferanten oder Investoren beeinflusst.

Abschließend wäre es auch interessant, die Rolle der so-

zialen Medien bei der Identifizierung von Greenwashing und der Mäßigung oder Verstärkung des Einflusses auf Stakeholder zu untersuchen. Frühe Arbeiten in diesem Bereich gehen davon aus, dass soziale Medien denjenigen Unternehmen zugutekommen, die an CSR-Aktivitäten teilnehmen und ehrlich über ihre Handlungen kommunizieren (vgl. K. Lee, Oh & Kim, 2013, S. 804). Weiter besteht die Hoffnung, dass Greenwashing im Zuge der fortschreitenden Digitalisierung zumindest verringert werden kann (vgl. F. Bowen, 2014, S. 29). Zahlreiche Websites machen inzwischen auf Greenwashing aufmerksam. Greenpeace hat bspw. das Green Marketing von Shell über die Website „Arctic Ready“ (www.arcticready.com) ins Visier genommen, und die Marketingbemühungen des Unternehmens angekreidet (vgl. Fernando, Suganthi & Sivakumaran, 2014, S. 169). Mithilfe sozialer Medien konnte „Arctic Ready“ extrem schnell verbreitet werden und viele Social Media Nutzer haben sich gegen Shell gewandt. Wenn genug Konsumenten oder Investoren reagieren, wäre es möglich, dass Greenwashing zu einem extremen Risiko für Unternehmen und somit unrentabel wird. (vgl. Lyon & Montgomery, 2013, S. 755). Eine Gefahr bei diesem Ansatz zur Verhinderung von Greenwashing besteht darin, dass auch tatsächlich nachhaltige Unternehmen ihr Green Marketing einschränken könnten, aus Angst, fälschlicherweise ins Visier von Organisationen wie Greenpeace zu geraten (vgl. Lyon & Maxwell, 2011, S. 29) oder davon abgehalten werden, überhaupt Verbesserungen vorzunehmen, was ebenfalls unvorteilhaft für die Gesellschaft wäre (vgl. Lyon & Montgomery, 2015, S. 240). Auch in diesem Bereich existieren demnach noch einige Forschungslücken. Um die Auswirkungen von sozialen Medien auf Greenwashing zu erforschen sollten Studien über Variationen zwischen den Kommunikationsmedien sowie verschiedenen Stakeholdern, Greenwashing-Methoden und Unternehmenscharakteristika durchgeführt werden.

Kurzum liefert diese Studie vielversprechende Ansätze, auf denen in der Marketingforschung zukünftig aufgebaut werden kann. Zu ergründen, wie genau CSR aufgebaut sein muss, damit es auch als solches wahrgenommen wird und eine ökologische Motivation erkennbar ist, kann dabei als zentrale Herausforderung extrahiert werden.

5.3. Limitationen

Wie jede empirische Studie weist auch diese Limitationen auf, die im folgenden Abschnitt dargelegt werden. Einige Einschränkungen sind bereits im vorigen Teil sichtbar geworden. So werden hier lediglich zwei verschiedene CSR-Initiativen, Greenwashing-Methoden und Greenwashing-Skandale untersucht. Es ist unklar, ob sich bei anderen Methoden ähnliche Ergebnisse zeigen. Außerdem wird hier lediglich Claim-Greenwashing betrachtet. Die Verwendung von naturanregenden Elementen, das sogenannte Executional Greenwashing, wird nicht untersucht.

Darüber hinaus werden die Kommunikationen lediglich in Kombination mit einer Marke und einem Produkt betrachtet. Adidas ist eine Marke mit guter Reputation, vor allem in Deutschland. Werden die Anzeigen bzw. Skandale mit einem anderen Unternehmen kombiniert, können sich ganz unterschiedliche Ergebnisse zeigen. Insbesondere wenn das Unternehmen in einer umweltsensiblen Branche tätig ist, sind stärkere Reaktionen auf Greenwashing und Greenwashing-Skandale zu erwarten (vgl. Cho et al., 2009, S. 949; Torelli et al., 2020, S. 416). Auch, dass einige Probanden möglicherweise sehr gut über die Umweltperformance von Adidas informiert sind, kann als Limitation aufgezeigt werden. Hier ist lediglich die generelle Einstellung der Probanden der Marke Adidas gegenüber abgefragt worden. Anders als bei der Untersuchung fiktiver Unternehmen kann hier nicht ausgeschlossen werden, dass einzelne Probanden Vorkenntnisse über die tatsächliche Umweltleistung von Adidas haben. Dadurch könnten die in dem Experiment gezeigten Werbeanzeigen unterschiedlich aufgenommen worden sein.

Weiterhin wird ausschließlich die Sicht der Konsumenten betrachtet. Die Effekte auf die Reputation aus Sicht anderer relevanter Stakeholder wie Mitarbeiter, Lieferanten, Investoren oder politische Entscheidungsträger werden nicht untersucht. Zudem ist die Stichprobe geprägt von Studenten sowie jungen Arbeitnehmern und spiegelt nicht die Grundgesamtheit wider (vgl. Tab. 3). Für repräsentative Ergebnisse fehlt ein größerer Anteil an älteren Personen sowie Personen mit höherem Einkommen.

Als Kommunikationsmedium für den Stimulus ist hier eine einfache Werbeanzeige mit zugehörigem Text verwendet worden. Green Marketing kann allerdings auch über andere Wege, wie Social Media Posts, TV-Spots oder Produktverpackungen kommuniziert werden. Die Ergebnisse können daher nicht für alle Medien pauschalisiert werden. Auch ein Greenwashing-Skandal hätte womöglich stärkere Effekte, wenn darüber im TV berichtet wird. Da hier aber ein Feldexperiment in Form einer Onlinebefragung konzipiert worden ist, konnte hier lediglich ein selbst verfasster Text in Form eines Zeitungsartikels präsentiert werden. Onlinebefragungen sind generell nicht frei von Restriktionen. Beispiele dafür sind Ablenkungen, die während der Beantwortung auftreten können sowie die Anonymität der Probanden (vgl. Wright, 2005, o. S.). In dieser Studie muss unter anderem von falschen Selbsteinschätzungen, bspw. bei den Fragen zum Umweltbewusstsein, ausgegangen werden.

Eine weitere Einschränkung dieser Untersuchung ist, dass es nicht möglich ist, die Handlungen und Verhaltensweisen der Teilnehmer nach dem Experiment zu überprüfen oder zu beurteilen, ob ihre Wahrnehmungen und Absichten nachhaltig beeinflusst wurden. Die Absicht zu haben, etwas zu tun, bedeutet nicht automatisch, dass es auch getan wird. Eine Kaufabsicht kann bspw. nicht mit der tatsächlichen Kaufhandlung gleichgesetzt werden.

Außerdem bleiben auch Einflüsse unbeobachtet, die den Akt des Kaufs betreffen. Ein Beispiel dafür ist der sogenannte konkurrierende Altruismus. Die Theorie des konkurrierenden Altruismus ist eine weitere Erklärung, warum so viele Verbraucher nachhaltige Produkte und Dienstleistungen nachfragen. Es ist ein soziales Phänomen, das definiert wird als ein Prozess, durch den Individuen versuchen, sich gegenseitig in Bezug auf ihre Großzügigkeit zu übertreffen (vgl. Hardy &

Van Vugt, 2006, S. 1403). Im Wesentlichen geht die Theorie davon aus, dass Individuen aus egoistischen Gründen versuchen, als altruistisch wahrgenommen zu werden, um den eigenen Status zu erhöhen. Demnach sind Konsumenten eher dazu bereit sich auf altruistische Handlungen einzulassen, also bspw. Bio-zertifizierte Produkte zu einem höheren Preis zu kaufen, wenn die Handlung beobachtet und dementsprechend erhöhte Anerkennung in der Gesellschaft erhofft wird (vgl. Hardy & Van Vugt, 2006, S. 1403). Im Rahmen einer anonymen Onlinebefragung können solche Phänomene nicht untersucht werden. Aus diesem Grund wäre es für zukünftige Forschung sinnvoll, die hier erzielten Ergebnisse empirisch zu validieren und ggf. reale Einkaufssituationen nachzustellen.

Generell sollten die Ergebnisse mittels Studien, bei denen andere Marken, Produkte, Labels, Probanden und Medien betrachtet werden, verifiziert werden. In dem Kontext fehlt es auch an längerfristigen Daten. Die hier durchgeführte Studie betrachtet lediglich die kurzfristigen Reaktionen auf nur eine Werbeanzeige, nicht aber die Effekte von jahrelangem Greenwashing über mehrere Kommunikationskanäle.

Zusammengefasst hat diese Studie zu durchaus nennenswerten Erkenntnissen geführt. Aufgrund der erläuterten Limitationen bleiben jedoch weitere Untersuchungsfelder offen, die es für ein Gesamtbild des Forschungsfeldes zu berücksichtigen gilt.

6. Fazit und Ausblick

Das Ziel dieser Arbeit war es, die Effekte nachhaltiger Unternehmenspraktiken und unterschiedlicher Greenwashing-Methoden auf die Unternehmensreputation zu untersuchen. Dazu sind zunächst die wichtigsten Begriffe in diesem Kontext herausgearbeitet und erläutert worden. Neben dem CSR sind auch Green Marketing, Greenwashing sowie die Unternehmensreputation definiert worden. Auf dieser Grundlage konnte ein Experiment erstellt und durchgeführt werden, das anhand verschiedener selbsterstellter Adidas-Werbeanzeigen neue nennenswerte Erkenntnisse in Bezug auf die Wirkungen unterschiedlicher „grüner“ Unternehmenskommunikationen liefert. Sowohl für CSR als auch für Greenwashing und die Skandale sind dafür die Methoden „Fake-Label“, bzw. „Umweltlabel“ sowie „Vagheit“, bzw. „Keine Vagheit“ verwendet worden.

Ein erstes zentrales Ergebnis ist, dass keine signifikanten Unterschiede zwischen den Effekten der jeweiligen Methoden einer Übergruppe festgestellt werden können. Bspw. sind die Effekte der Greenwashing-Methode, bei der ein Fake-Label verwendet wurde, ähnlich zu der mit vagen/irreführenden Aussagen.

Weiterhin sind auch keine signifikanten Unterschiede zwischen der Kontrollgruppe, den CSR-Initiativen sowie den Greenwashing-Methoden zu erkennen. Lediglich die Greenwashing-Anzeigen mit nachgelagerten Zeitungsartikeln, in denen über einen Greenwashing-Skandal bei Adidas berichtet wird, haben zu nennenswerten negativen Effekten auf die Unternehmensreputation geführt.

Als Begründungen für diese Ergebnisse können insbesondere die hohe Skepsis und die Verwirrung der Probanden herausgearbeitet werden. In allen vier Experimentalgruppen, in denen die Probanden nicht darüber informiert worden sind, ob es sich bei der Werbeanzeige um tatsächliches CSR oder Greenwashing handelt, ist Greenwashing relativ stark wahrgenommen worden (vgl. Tab. 9). Dabei sind CSR-Initiativen häufig als Greenwashing und Greenwashing-Methoden als tatsächliches CSR wahrgenommen worden. Ist bei den CSR-Initiativen Greenwashing sogar stark wahrgenommen worden, wurde die Reputation von Adidas schlechter bewertet als in der Kontrollgruppe. Bei den Greenwashing-Methoden zeigen sich umgekehrt die gleichen Effekte (vgl. Tab. 11 & 12). Es existiert demnach eine starke Verwirrung bei den Probanden, weshalb die CSR-Initiativen und Greenwashing-Methoden häufig nicht als solche erkannt werden können.

Bei der Betrachtung des Strukturmodells zeigt sich weiterhin, dass die Wahrnehmung von Greenwashing anders als auf die RepTrak-Faktoren nur sehr geringe, nicht signifikante Effekte auf die Markenwahrnehmung und das Markenverhalten hat (vgl. Abb. 6). Die Reputation scheint in der kurzfristigen Perspektive die negativen Wirkungen auf das Verhalten der Konsumenten abzumildern.

Darüber hinaus ist die wahrgenommene ökologische/soziale Motivation der Unternehmen entscheidend für den Erfolg von CSR-Initiativen. Während sich herausstellt, dass eine strategische Motivation akzeptiert wird, ohne einen Treiber für die Effekte auf die Reputation darzustellen, sind bei der ökologischen Motivation deutlich stärkere Auswirkungen erkennbar. Nehmen die Konsumenten eine starke ökologische Motivation für die nachhaltige Unternehmensaktivität wahr, werden sowohl nahezu alle RepTrak-Dimensionen als auch die Markenwahrnehmung und das Markenverhalten signifikant besser bewertet (vgl. Tab. 14). Auch hier sind die Probanden allerdings häufig nicht dazu in der Lage gewesen, zwischen CSR und Greenwashing zu unterscheiden (vgl. Anhang Tab. 30).

Greenwashing ist als zentraler Grund für die Skepsis und Verwirrung der Konsumenten hervorzuheben (vgl. Jahdi & Acikdilli, 2009, S. 111), wobei dessen Ausbreitung sowohl die Effektivität der tatsächlichen CSR-Politik (vgl. Elving, Van Vuuren et al., 2011, S. 49-55) als auch die globale Entwicklung nachhaltigerer Unternehmen (vgl. Alves, 2009, S. 2-3) bedroht. Es wird deutlich, dass alle von einer Reduzierung des Greenwashings profitieren: (1) Konsumenten würden wieder einen Sinn hinter Umweltlabels erkennen, statt den Wahrheitsgehalt aller Umweltbehauptungen in Frage zu stellen; (2) Unternehmen, die sich entschieden um Nachhaltigkeit bemühen, würden nicht aufgrund von Verallgemeinerungen bestraft, die Konsumenten aufgrund von Verfehlungen der Konkurrenz machen; und (3) die Umwelt würde stärker von den gemeinsamen Bemühungen von Unternehmen und Konsumenten profitieren (vgl. L. Mitchell & Ramey, 2011, S. 43).

Aktuell trägt das niedrige Umweltwissen (EK) der Konsumenten allerdings noch nicht zu einer solchen Entwicklung bei. Von den insgesamt 517 Probanden in dieser Studie ha-

ben lediglich 116, also etwa jeder fünfte, erkennen können, ob es sich bei den vier gezeigten Symbolen um zertifizierte Umweltlabel oder Zeichen ohne jegliche Aussagekraft handelt (vgl. Tab. 23). Solange Konsumenten die Bedeutung eines Symbols nicht kennen, das auf einer Werbeanzeige oder Verpackung abgebildet ist, führt dies konsequenterweise zu Skepsis sowie Verwirrung und es ist schwieriger, Greenwashing von CSR zu unterscheiden.

Um dem entgegenzuwirken könnten Regierungsinitiativen und Sensibilisierungskampagnen das Wissen der Konsumenten erweitern, damit ein besseres Verständnis von Umweltaussagen in der Werbung gewährleistet werden kann. Zu diesem Zweck könnten zusätzliche Informationen, bspw. über Vergleichstools (z.B. Nutri-Score), QR-Codes oder Apps nützlich sein, die den Konsumenten helfen, CSR und Greenwashing eher zu erkennen.

Weiterhin könnten neue, nur wenige, dafür sehr transparente Umweltlabel von Politik, Wirtschaft und unabhängigen Experten entworfen werden. Aktuell ist die hohe Anzahl verschiedener Labels sehr verwirrend für Konsumenten. Die Kriterien sind undurchsichtig und Fake-Labels werden immer häufiger (vgl. TerraChoice, 2010, S. 19). Bei einer geringen Anzahl transparenter Labels, die den Konsumenten bekannt sind und denen sie vertrauen können, ließen sich tatsächlich nachhaltige Produkte deutlich einfacher erkennen und die Skepsis und Verwirrung der Konsumenten würde sich verringern. Dabei wird häufig angenommen, dass sich die Industrie gegen Umweltzertifizierungen sträubt, obwohl dies oft nicht der Fall ist. In mehreren Geschäftsbereichen gibt es viele Unternehmen, die sich für die Einführung von branchenweiten Verhaltenskodizes oder Regeln für Produktkategorien einsetzen. Dazu gehören bspw. die Tourismusbranche (vgl. V. L. Smith & Font, 2014, S. 942), Versicherungen (vgl. Mills, 2009, S. 323) oder auch Öl- und Gasunternehmen (vgl. Stephenson, Doukas & Shaw, 2012, S. 458). Dabei ist eine Offenlegung der Kriterien bedeutsam, die für die Vergabe des Labels zu erfüllen sind, sowie die stetige Überprüfung der zertifizierten Unternehmen durch unabhängige Organisationen. Ebendies sind Punkte, die aktuell selbst bei weit verbreiteten Labels wie des Marine Steward-ship Council (MSC) kritisiert werden. Umweltorganisationen wie bspw. Greenpeace bemängeln hier unter anderem Interessenkonflikte mit der Industrie, da ein Großteil der Einnahmen durch Zertifizierungsgebühren der Unternehmen zustande kommt, sodass MSC einen starken Anreiz hat, viele Unternehmen zu zertifizieren (vgl. Greenpeace, 2021).

Allerdings kann davon ausgegangen werden, dass auch solche Labels Greenwashing nicht vollständig abwehren können. Executional Greenwashing, also die Verwendung naturanregender Elemente oder auch andere Greenwashing-Sünden wie die des versteckten Kompromisses, werden weiterhin kaum zu verhindern sein. Letzteres tritt auf, wenn Unternehmen nur selektiv positive Eigenschaften offenlegen, während sie negative Auswirkungen ignorieren (vgl. TerraChoice, 2010, S. 10).

Eine weitere begrenzte, aber potenziell vielversprechende Möglichkeit zur Reduzierung von Skepsis könnte auf der Ebene der Einzelhändler liegen. Laut der Theorie des konkurrierenden Altruismus achten Konsumenten insbesondere in öffentlichen Einkaufssituationen auf die Nachhaltigkeit von Produkten (vgl. Hardy & Van Vugt, 2006, S. 1403). Erfahrene und informierte Einzelhändler könnten daher ggf. tatsächlich nachhaltige Produkte hervorheben oder Informationstafeln zu verschiedenen Umweltlabels einführen (vgl. Cliath, 2007, S. 434-435).

Zudem könnte das höhere Risiko, das für unehrliche Unternehmen aufgrund von Social-Media aufgekommen ist, Greenwashing weiter verringern und folglich das Vertrauen in CSR erhöhen (vgl. Lyon & Montgomery, 2013, S. 755).

Die Absicht dieser Arbeit ist es weder, von Green Marketing abzuraten, noch sollen Konsumenten von nachhaltigen Werbeanzeigen abgeschreckt werden. Vielmehr sollen mit dieser Studie Marketingverantwortliche dabei unterstützt werden, effektive sowie aufrichtige Unternehmenspraktiken aufzubauen und zu kommunizieren, um die Herausforderungen durch Greenwashing auf dem Markt anzugehen. Dadurch wären die Konsumenten in der Lage, Green Marketing zu vertrauen und nachhaltige Produkte könnten sich schneller auf dem Markt durchsetzen (vgl. TerraChoice, 2010, S. 6). Die hohe Skepsis und Verwirrung stellen aktuell noch ein großes Problem dar. Um dieses zu lösen, müssen die Quellen dieser Skepsis bzw. Verwirrung systematisch beseitigt werden, wofür insbesondere Greenwashing unterbunden werden muss. Das kann nur gelingen, wenn nachhaltige Unternehmen, die Politik, unabhängige Experten, Organisationen und Einzelhändler zusammenarbeiten.

Zusammenfassend wird diese Studie die akademische Forschung von CSR, Greenwashing und Umweltverantwortung bereichern und eine Forschungslücke in der Literatur über die Effekte von CSR-Initiativen, Greenwashing-Methoden und Greenwashing-Skandalen auf die Reputation von Unternehmen schließen.

Die Zukunft von CSR wird die neuesten technologischen Fortschritte und ihre Rolle als Teil neuer Geschäftsprozesse berücksichtigen müssen. Die Übernahme und Anpassung an neue Digitalisierungsprozesse und -tools sowie die Einbindung von künstlicher Intelligenz in das Geschäftsumfeld sind relevante Herausforderungen, sowohl für die CSR-Debatte als auch für Unternehmen im Allgemeinen. Gleichzeitig bieten sich darüber hinaus eine Vielzahl von Chancen, um die Gesellschaft ein wenig gemeinnütziger zu gestalten. Die Welt steht mit dem Klimawandel vor einem Problem unüberschaubaren Ausmaßes, das ohne ein nachhaltigeres Bewusstsein sowohl der Konsumenten als auch der Unternehmen weiter zunehmen wird. Um dem entgegenzutreten, benötigt es Zusammenarbeit sowie die Motivation, gemeinsame Werte zu schaffen – und genau das ist die Grundidee von Corporate Social Responsibility.

Literatur

Aaron, J. R., McMillan, A. & Cline, B. N. (2012). Investor reaction to firm environmental management reputation. *Corporate Reputation Review*, *15* (4), 304–318.

Adidas. (2019). *Adidas geschäftsbericht 2019*. Zugriff auf https://report.adidas-group.com/2019/de/serviceseiten/downloads/files/adidas_geschaeftsbericht_2019.pdf (abgerufen am 10.04.2021)

Aguilera, R. V. & Jackson, G. (2010). Comparative and international corporate governance. *Academy of Management annals*, *4* (1), 485–556.

Aguinis, H. & Glavas, A. (2012). What we know and don't know about corporate social responsibility: A review and research agenda. *Journal of management*, *38* (4), 932–968.

Aji, H. M. & Sutikno, B. (2015). The extended consequence of greenwashing: Perceived consumer skepticism. *International Journal of Business and Information*, *10* (4), 433–468.

Akinwande, M. O., Dikko, H. G., Samson, A. et al. (2015). Variance inflation factor: as a condition for the inclusion of suppressor variable (s) in regression analysis. *Open Journal of Statistics*, *5* (07), 754.

Alexander, E. C. (2002). Consumer reactions to unethical service recovery. *Journal of business Ethics*, *36* (3), 223–237.

Alniacik, E., Alniacik, U. & Erdogmus, N. (2012). How do the dimensions of corporate reputation affect employment intentions? *Corporate Reputation Review*, *15* (1), 3–19.

Alves, I. M. (2009). Green spin everywhere: How greenwashing reveals the limits of the csr paradigm. *Journal of Global Change & Governance*, *2* (1), 1–26.

Andrews, J. C., Burton, S. & Netemeyer, R. G. (2000). Are some comparative nutrition claims misleading? the role of nutrition knowledge, ad claim type and disclosure conditions. *Journal of Advertising*, *29* (3), 29–42.

Anuar, M. M., Omar, K. & Mohamad, O. (2013). Does skepticism influence consumers intention to purchase cause-related products? *International Journal of Business and Social Science*, *4* (5).

Argenti, P. A. & Druckenmiller, B. (2004). Reputation and the corporate brand. *Corporate reputation review*, *6* (4), 368–374.

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2016). *Multivariate analysemethoden*. Berlin, Heidelberg: Springer.

Bagozzi, R. P. & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the academy of marketing science*, *16* (1), 74–94.

Bansal, P. & Roth, K. (2000). Why companies go green: A model of ecological responsiveness. *Academy of management journal*, *43* (4), 717–736.

Baron, D. P. (2001). Private politics, corporate social responsibility, and integrated strategy. *Journal of economics & management strategy*, *10* (1), 7–45.

Basdeo, D. K., Smith, K. G., Grimm, C. M., Rindova, V. P. & Derfus, P. J. (2006). The impact of market actions on firm reputation. *Strategic Management Journal*, *27* (12), 1205–1219.

Baum, L. M. (2012). It's not easy being green or is it? a content analysis of environmental claims in magazine advertisements from the united states and united kingdom. *Environmental Communication: A Journal of Nature and Culture*, *6* (4), 423–440.

Baumgartner, R. J. (2014). Managing corporate sustainability and csr: A conceptual framework combining values, strategies and instruments contributing to sustainable development. *Corporate Social Responsibility and Environmental Management*, *21* (5), 258–271.

Bendisch, F., Larsen, G. & Trueman, M. (2013). Fame and fortune: a conceptual model of ceo brands. *European journal of marketing*, *47* (3/4), 596–614.

Besel, R. D., Burke, K. & Christos, V. (2017). A life history approach to perceptions of global climate change risk: young adults' experiences about impacts, causes, and solutions. *Journal of Risk Research*, *20* (1), 61–75.

Bhatia, M. & Jain, A. (2013). Green marketing: A study of consumer perception and preferences in india. *Electronic Green Journal*, *1* (36).

Bickart, B. A. & Ruth, J. A. (2012). Green eco-seals and advertising persuasion. *Journal of advertising*, *41* (4), 51–67.

Bortz, J. (2005). *Statistik für human-und sozialwissenschaftler (2010)*. 6. Auflage, Heidelberg: Springer-Verlag.

Bowen, F. (2014). *After greenwashing: Symbolic corporate environmentalism and society*. Cambridge University Press.

Bowen, H. R. (1953). *Social responsibilities of the businessman*. University of Iowa Press.

Bowie, N. E. (2017). *Business ethics: A kantian perspective*. Cambridge: Cambridge University Press.

Brosius, F. (2006). *SPSS 14: das mitp-Standardwerk; fundierte Einführung in SPSS und die Statistik–alle statistischen Verfahren mit praxisnahen Beispielen*. Heidelberg: mitp.

Brown, T. J. & Dacin, P. A. (1997). The company and the product: Corporate associations and consumer product responses. *Journal of marketing*, *61* (1), 68–84.

Buck, R., Anderson, E., Chaudhuri, A. & Ray, I. (2004). Emotion and reason in persuasion: Applying the ari model and the casc scale. *Journal of Business Research*, *57* (6), 647–656.

Bühner, M. (2010). *Einführung in die Test- und Fragebogenkonstruktion. 2., aktualisierte und erweiterte Auflage*. Mnchen: Pearson Studium.

BUND. (2016). *Greenwash bei british petroleum – der bp-schwindel.* Zugriff auf http://www.bund-rvso.de/bp-schwindel.html (abgerufen am 10.04.2021)

Burke, L. & Logsdon, J. M. (1996). How corporate social responsibility pays off. *Long range planning*, *29* (4), 495–502.

Burns, A. C., Biswas, A. & Babin, L. A. (1993). The operation of visual imagery as a mediator of advertising effects. *Journal of Advertising*, *22* (2), 71–85.

Carmeli, A. & Freund, A. (2002). The relationship between work and workplace attitudes and perceived external prestige. *Corporate reputation review*, *5* (1), 51–68.

Carmines, E. G. & Zeller, R. A. (1979). *Reliability and validity assessment*. Newbury Park, CA: Sage publications.

Carroll, A. B. (1979). A three-dimensional conceptual model of corporate performance. *Academy of management review*, *4* (4), 497–505.

Carroll, A. B. (2008). A history of corporate social responsibility: Concepts and practices. *The Oxford handbook of corporate social responsibility*, 19–46.

Carroll, A. B. (2015). Corporate social responsibility: The centerpiece of competing and complementary frameworks. *Organizational dynamics*.

Carroll, A. B. et al. (1991). The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders. *Business horizons*, *34* (4), 39–48.

Casado, A. M., Peláez, J. I. & Cardona, J. (2014). Managing corporate reputation: A perspective on the spanish market. *Corporate Reputation Review*, *17* (1), 46–63.

Certo, S. T. (2003). Influencing initial public offering investors with prestige: Signaling with board structures. *Academy of management review*, *28* (3), 432–446.

Chan, S. Y. & Leung, P. (2006). The effects of accounting students' ethical reasoning and personal factors on their ethical sensitivity. *Managerial auditing journal*.

Chandler, D. (2016). *Strategic corporate social responsibility: Sustainable value creation*. Sage Publications.

Chen, X. & Kelly, T. F. (2015). B-corps–a growing form of social enterprise: Tracing their progress and assessing their performance. *Journal of Leadership & Organizational Studies*, *22* (1), 102–114.

Chen, Y.-S. (2008). The driver of green innovation and green image–green core competence. *Journal of business ethics*, *81* (3), 531–543.

Chen, Y.-S. & Chang, C.-H. (2013). Greenwash and green trust: The mediation effects of green consumer confusion and green perceived risk. *Journal of business ethics*, *114* (3), 489–500.

Chen, Y.-S., Lin, C.-L. & Chang, C.-H. (2014). The influence of greenwash on green word-of-mouth (green wom): The mediation effects of green perceived quality and green satisfaction. *Quality & Quantity*, *48* (5), 2411–2425.

Chin, W. W. (2010). How to write up and report pls analyses. In *Handbook of partial least squares* (S. 655–690). Springer.

Chin, W. W. et al. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research*, *295* (2), 295–336.

Cho, C. H., Phillips, J. R., Hageman, A. M. & Patten, D. M. (2009). Media richness, user trust, and perceptions of corporate social responsibility: An experimental investigation of visual web site disclosures. *Accounting, Auditing & Accountability Journal*, *22* (6), 933–952.

Cleff, T. (2015). *Faktorenanalyse. in deskriptive statistik und explorative datenanalyse*. Wiesbaden: Gabler Verlag.

Cliath, A. G. (2007). Seeing shades: Ecological and socially just labeling. *Organization & Environment*, *20* (4), 413–439.

Concise Oxford English Dictionary. (2010). *10th edition*. Oxford: Oxford University Press.

Connelly, B. L., Certo, S. T., Ireland, R. D. & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of management*, *37* (1), 39–67.

Connors, S., Anderson-MacDonald, S. & Thomson, M. (2017). Overcoming the 'window dressing' effect: Mitigating the negative effects of inherent skepticism towards corporate social responsibility. *Journal of business ethics*, *145* (3), 599–621.

Costa, R. & Menichini, T. (2013). A multidimensional approach for csr assessment: The importance of the stakeholder perception. *Expert systems with applications*, *40* (1), 150–161.

Cottan-Nir, O. (2019). Toward a conceptual model for determining ceo brand equity. *Corporate Reputation Review*, *22* (4), 121–133.

Courtright, J. L. & Smudde, P. M. (2009). Leveraging organizational innovation for strategic reputation management. *Corporate Reputation Review*, *12* (3), 245–269.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, *16* (3), 297–334.

Cullen, J. B., Parboteeah, K. P. & Victor, B. (2003). The effects of ethical climates on organizational commitment: A two-study analysis. *Journal of business ethics*, *46* (2), 127–141.

Davis, G. F. (2005). New directions in corporate governance. *Annual review of sociology*, 143–162.

Davis, K. (1960). Can business afford to ignore social responsibilities? *California management review*, *2* (3), 70–76.

Dawar, N. & Parker, P. (1994). Marketing universals: Consumers' use of brand name, price, physical appearance, and retailer reputation as signals of product quality. *Journal of marketing*, *58* (2), 81–95.

De Jong, M. D., Harkink, K. M. & Barth, S. (2018). Making green stuff? effects of corporate greenwashing on consumers. *Journal of business and technical communication*, *32* (1), 77–112.

Delmas, M. A. & Burbano, V. C. (2011). The drivers of greenwashing. *California management review*, *54* (1), 64–87.

Delmas, M. A. & Toffel, M. W. (2008). Organizational responses to environmental demands: Opening the black box. *Strategic management journal*, *29* (10), 1027–1055.

Dembek, K., York, J. & Singh, P. J. (2018). Creating value for multiple stakeholders: Sustainable business models at the base of the pyramid. *Journal of Cleaner production*, *196*, 1600–1612.

De Vries, G., Terwel, B. W., Ellemers, N. & Daamen, D. D. (2015). Sustainability or profitability? how communicated motives for environmental policy affect public perceptions of corporate greenwashing. *Corporate Social Responsibility and Environmental Management*, *22* (3), 142–154.

Divine, R. L. & Lepisto, L. (2005). Analysis of the healthy lifestyle consumer. *Journal of Consumer marketing*.

Do Paço, A. M. F. & Reis, R. (2012). Factors affecting skepticism toward green advertising. *Journal of advertising*, *41* (4), 147–155.

Dowling, G. R. & Gardberg, N. A. (2012). Keeping score: The challenges of measuring corporate reputation. *The Oxford handbook of corporate reputation*, 34–68.

Du, S., Bhattacharya, C. B. & Sen, S. (2007). Reaping relational rewards from corporate social responsibility: The role of competitive positioning. *International journal of research in marketing*, *24* (3), 224–241.

Du, S., Bhattacharya, C. B. & Sen, S. (2010). Maximizing business returns to corporate social responsibility (csr): The role of csr communication. *International journal of management reviews*, *12* (1), 8–19.

Du, X. (2015). How the market values greenwashing? evidence from china. *Journal of Business Ethics*, *128* (3), 547–574.

Duffy, M. K., Ganster, D. C. & Pagon, M. (2002). Social undermining in the workplace. *Academy of management Journal*, *45* (2), 331–351.

Eberl, M. & Schwaiger, M. (2005). Corporate reputation: disentangling the effects on financial performance. *European Journal of marketing*, *39* (7/8), 838–854.

Elkington, J. (1998). Partnerships from cannibals with forks: The triple bottom line of 21st-century business. *Environmental quality management*, *8* (1), 37–51.

Ellen, P. S. (1994). Do we know what we need to know? objective and subjective knowledge effects on pro-ecological behaviors. *Journal of Business Research*, *30* (1), 43–52.

Elving, W., Van Vuuren, M. et al. (2011). Beyond identity washing: Corporate social responsibility in an age of skepticism. *Akademija Mm*, *17*, 49–55.

Enobong, U. (2017). Implications of accountants unethical behavior and corporate failures. *International Journal of Business, Economics and Management*, *4* (4), 82–94.

Esch, F.-R., Langner, T. & Brunner, C. (2005). Kundenbezogene ansätze des markencontrolling. In *Moderne markenführung* (S. 1227–1261). Springer.

Evan, W. M. & Freeman, R. E. (1988). *A stakeholder theory of the modern corporation: Kantian capitalism*. (in T.L. Beauchamp and N. Bowie (eds.), Ethical Theory and Business (Englewood Cliffs, NJ: Prentice Hall): 75-84)

Faircloth, J. B., Capella, L. M. & Alford, B. L. (2001). The effect of brand attitude and brand image on brand equity. *Journal of marketing theory and practice*, *9* (3), 61–75.

Fan, D., Geddes, D. & Flory, F. (2013). The toyota recall crisis: Media impact on toyota's corporate brand reputation. *Corporate Reputation Review*, *16* (2), 99–117.

Fang, E., Palmatier, R. W. & Grewal, R. (2011). Effects of customer and innovation asset configuration strategies on firm performance. *Journal of marketing research*, *48* (3), 587–602.

Federal Trade Commission. (2012). Guides for the use of environmental marketing claims; final rule. *Federal Register*, *77* (197), 62122–62132.

Feinstein, N. (2013). Learning from past mistakes: Future regulation to prevent greenwashing. *Boston College Environmental Affairs Law Review*, *40* (1), 229–258.

Fernando, A. G., Suganthi, L. & Sivakumaran, B. (2014). If you blog, will they follow? using online media to set the agenda for consumer concerns on "greenwashed" environmental claims. *Journal of Advertising*, *43* (2), 167–180.

Filatotchev, I. & Bishop, K. (2002). Board composition, share ownership, and "underpricing" of uk ipo firms. *Strategic management journal*, *23* (10), 941–955.

Flatt, S. J., Harris-Boundy, J. & Wagner, S. (2013). Ceo succession: A help or hindrance to corporate reputation? *Corporate Reputation Review*, *16* (3), 206–219.

Fombrun, C. & Shanley, M. (1990). What's in a name? reputation building and corporate strategy. *Academy of management Journal*, *33* (2), 233–258.

Fombrun, C. J., Gardberg, N. A. & Sever, J. M. (2000). The reputation quotientsm: A multi-stakeholder measure of corporate reputation. *Journal of brand management*, *7* (4), 241–255.

Fombrun, C. J., Ponzi, L. J. & Newburry, W. (2015). Stakeholder tracking and analysis: The reptrak® system for measuring corporate reputation. *Corporate reputation review*, *18* (1), 3–24.

Fombrun, C. J. & Rindova, V. P. (2001). Fanning the flame: Corporate reputations as social constructions of performance. *Constructing Markets and Industries, Oxford University Press, New York*, *53* (4), 5–14.

Foreh, M. R. & Grier, S. (2003). When is honesty the best policy? the effect of stated company intent on consumer skepticism. *Journal of consumer psychology*, *13* (3), 349–356.

Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, *18* (1), 39–50.

Friedman, M. (1962). *Capitalism and freedom*. Chicago: University of Chicago Press.

Friedman, M. (1970). A friedman doctrine: The social responsibility of business is to increase its profits. *The New York Times Magazine*, *13* (1970), 32–33.

Furlow, N. E. (2010). Greenwashing in the new millennium. *The Journal of Applied Business and Economics*, *10* (6), 22–25.

Ghosh, M. & John, G. (2009). When should original equipment manufacturers use branded component contracts with suppliers? *Journal of Marketing Research*, *46* (5), 597–611.

Goldfarb, A. & Yang, B. (2009). Are all managers created equal? *Journal of*

Marketing research, *46* (5), 612–622.
Gottschalk, P. (2013). Limits to corporate social responsibility: the case of gjensidige insurance company and hells angels motorcycle club. *Corporate Reputation Review*, *16* (3), 177–186.
Graafland, J. & Van de Ven, B. (2006). Strategic and moral motivation for corporate social responsibility. *Journal of Corporate Citizenship* (22), 111–123.
Gray, E. R. & Balmer, J. M. (1998). Managing corporate image and corporate reputation. *Long Range Planning*, *31* (5), 695–702.
Greenpeace. (2021). *Greenpeace-position zum "marine stewardship council" (msc)*. Zugriff auf https://www.greenpeace.de/themen/meere/greenpeace-position-zum-marine-stewardship-council-msc (abgerufen am 10.04.2021)
Grimm, M., Högl, S. & Hupp, O. (2000). Target® positioning–ein bewährtes tool zur unterstützung des strategischen markenmanagements. *Jahrbuch der Absatz-und Verbrauchsforschung*, *1/2000*, 4–18.
Grimmer, M. & Bingham, T. (2013). Company environmental performance and consumer purchase intentions. *Journal of business research*, *66* (10), 1945–1953.
Hair, J. F., Ringle, C. M. & Sarstedt, M. (2011). Pls-sem: Indeed a silver bullet. *Journal of Marketing theory and Practice*, *19* (2), 139–152.
Halff, G. (2013). The presentation of ceos in economic downturn. *Corporate Reputation Review*, *16* (3), 234–243.
Hamann, R. & Kapelus, P. (2004). Corporate social responsibility in mining in southern africa: Fair accountability or just greenwash? *Development*, *47* (3), 85–92.
Hardy, C. L. & Van Vugt, M. (2006). Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin*, *32* (10), 1402–1413.
Hartmann, P., Apaolaza, V. & Alija, P. (2013). Nature imagery in advertising: Attention restoration and memory effects. *International Journal of Advertising*, *32* (2), 183–210.
Hartmann, P., Apaolaza, V. & Eisend, M. (2016). Nature imagery in non-green advertising: The effects of emotion, autobiographical memory, and consumer's green traits. *Journal of Advertising*, *45* (4), 427–440.
Hartmann, P. & Apaolaza-Ibáñez, V. (2009). Green advertising revisited: Conditioning virtual nature experiences. *International Journal of Advertising*, *28* (4), 715–739.
Hartmann, P. & Apaolaza-Ibáñez, V. (2012). Consumer attitude and purchase intention toward green energy brands: The roles of psychological benefits and environmental concern. *Journal of business Research*, *65* (9), 1254–1263.
Helweg-Larsen, M. & Shepperd, J. A. (2001). Do moderators of the optimistic bias affect personal or target risk estimates? a review of the literature. *Personality and social psychology review*, *5* (1), 74–95.
Henard, D. H. & Dacin, P. A. (2010). Reputation for product innovation: Its impact on consumers. *Journal of Product Innovation Management*, *27* (3), 321–335.
Hoffman, A. (2013). Sc johnson and the greenlist backlash. *Ann Arbor, MI: William Davidson Institute, University of Michigan*.
Höflinger, P. J., Nagel, C. & Sandner, P. (2018). Reputation for technological innovation: Does it actually cohere with innovative activity? *Journal of Innovation & Knowledge*, *3* (1), 26–39.
Högl, S. & Hupp, O. (2004). Brand performance measurement mit dem brand assessment system (bass). *Der Wert der Marke*, 124–145.
Homburg, C. & Baumgartner, H. (1995). Beurteilung von kausalmodellen: Bestandsaufnahme und anwendungsempfehlungen. *Marketing: Zeitschrift für Forschung und Praxis*, 162–176.
Homburg, C. & Giering, A. (1996). [Komplexe Konstrukte]: Konzeptualisierung und Operationalisierung komplexer Konstrukte. *Marketing - Zeitschrift für Forschung und Praxis*, *18., Heft I*, 5–24.
Hsu, T. (2011). *Skepticism grows over products touted as eco-friendly. los angeles times, 21*. Zugriff auf https://www.latimes.com/business/la-xpm-2011-may-21-la-fi-greenwash-20110521-story.html (abgerufen am 10.04.2021)
Huang, R. & Sarigöllü, E. (2014). How brand awareness relates to market outcome, brand equity, and the marketing mix. In *Fashion branding and consumer behaviors* (S. 113–132). Springer.
Hulland, J. (1999). Use of partial least squares (pls) in strategic management research: A review of four recent studies. *Strategic management journal*, *20* (2), 195–204.
Jahdi, K. S. & Acikdilli, G. (2009). Marketing communications and corporate social responsibility (csr): Marriage of convenience or shotgun wedding? *Journal of business ethics*, *88* (1), 103–113.
Javed, M., Rashid, M. A., Hussain, G. & Ali, H. Y. (2020). The effects of corporate social responsibility on corporate reputation and firm financial performance: Moderating role of responsible leadership. *Corporate Social Responsibility and Environmental Management*, *27* (3), 1395–1409.
Jermsittiparsert, K., Siam, M., Issa, M., Ahmed, U. & Pahi, M. (2019). Do consumers expect companies to be socially responsible? the impact of corporate social responsibility on buying behavior. *Uncertain Supply Chain Management*, *7* (4), 741–752.
Jo, H. & Harjoto, M. A. (2012). The causal effect of corporate governance on corporate social responsibility. *Journal of business ethics*, *106* (1), 53–72.
Johar, G. V. (1995). Consumer involvement and deception from implied advertising claims. *Journal of Marketing Research*, *32* (3), 267–279.
Jones, R. (2005). Finding sources of brand value: Developing a stakeholder model of brand equity. *Journal of brand management*, *13* (1), 10–32.
Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American economic review*, *93* (5), 1449–1475.
Kahneman, D. & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management science*, *39* (1), 17–31.
Kaiser, H. F. & Rice, J. (1974). Little jiffy, mark iv. *Educational and psychological measurement*, *34* (1), 111–117.
Kangun, N., Carlson, L. & Grove, S. J. (1991). Environmental advertising claims: A preliminary investigation. *Journal of public policy & marketing*, *10* (2), 47–58.
Kantabutra, S. & Avery, G. C. (2010). The power of vision: statements that resonate. *Journal of business strategy*, *31* (1), 37–45.
Keh, H. T. & Xie, Y. (2009). Corporate reputation and customer behavioral intentions: The roles of trust, identification and commitment. *Industrial marketing management*, *38* (7), 732–742.
Kim, J. & Lennon, S. J. (2013). Effects of reputation and website quality on online consumers' emotion, perceived risk and purchase intention: Based on the stimulus-organism-response model. *Journal of Research in Interactive Marketing*.
Kim, K.-H., Kim, M. & Qian, C. (2018). Effects of corporate social responsibility on corporate financial performance: A competitive-action perspective. *Journal of management*, *44* (3), 1097–1118.
Kim, S. K., McFarland, R. G., Kwon, S., Son, S. & Griffith, D. A. (2011). Understanding governance decisions in a partially integrated channel: A contingent alignment framework. *Journal of Marketing Research*, *48* (3), 603–616.
King, A. & Lenox, M. (2002). Exploring the locus of profitable pollution reduction. *Management Science*, *48* (2), 289–299.
King, A. A., Lenox, M. J. & Terlaak, A. (2005). The strategic use of decentralized institutions: Exploring certification with the iso 14001 management standard. *Academy of management journal*, *48* (6), 1091–1106.
Kirchhoff, S. (2000). Green business and blue angels. *Environmental and Resource Economics*, *15* (4), 403–420.
Kivimaa, P. & Kautto, P. (2010). Making or breaking environmental innovation? technological change and innovation markets in the pulp and paper industry. *Management Research Review*, *33* (4), 289–305.
Kok, R. A. & Biemans, W. G. (2009). Creating a market-oriented product innovation process: A contingency approach. *Technovation*, *29* (8), 517–526.
Kolk, A. (2016). The social responsibility of international business: From ethics and the environment to csr and sustainable development. *Journal of World Business*, *51* (1), 23–34.
Koschate-Fischer, N., Stefan, I. V. & Hoyer, W. D. (2012). Willingness to pay for cause-related marketing: The impact of donation amount and moderating effects. *Journal of marketing research*, *49* (6), 910–927.
Kuß, A., Wildner, R. & Kreis, H. (2014). *Marktforschung: Grundlagen der datenerhebung und datenanalyse*. Springer-Verlag.
Lane, E. L. (2010). Consumer protection in the eco-mark era: A preliminary survey and assessment of anti-greenwashing activity and eco-mark enforcement. *John Marshall Review of Intellectual Property Law*, *9* (3), 742–773.
Lange, D., Lee, P. M. & Dai, Y. (2011). Organizational reputation: A review.

Journal of management, *37* (1), 153–184.

Latapí Agudelo, M. A., Jóhannsdóttir, L. & Davídsdóttir, B. (2019). A literature review of the history and evolution of corporate social responsibility. *International Journal of Corporate Social Responsibility*, *4* (1), 1–23.

Laufer, W. S. (2003). Social accountability and corporate greenwashing. *Journal of business ethics*, *43* (3), 253–261.

Lee, K., Oh, W.-Y. & Kim, N. (2013). Social media for socially responsible firms: Analysis of fortune 500's twitter profiles and their csr/csir ratings. *Journal of business ethics*, *118* (4), 791–806.

Lee, S., Lee, K., Gao, Y. L., Xiao, Q. & Conklin, M. (2018). Do a company's sincere intentions with csr initiatives matter to employees? a comparison of customer-related and employee-related csr initiatives. *Journal of Global Responsibility*.

Lewis, S. (2003). Reputation and corporate responsibility. *Journal of Communication Management*, *7* (4), 356–366.

Lohmöller, J.-B. (1989). Predictive vs. structural modeling: Pls vs. ml. In *Latent variable path modeling with partial least squares* (S. 199–226). Springer.

Lyon, T. P. & Maxwell, J. W. (2008). Corporate social responsibility and the environment: A theoretical perspective. *Review of environmental economics and policy*, *2* (2), 240–260.

Lyon, T. P. & Maxwell, J. W. (2011). Greenwash: Corporate environmental disclosure under threat of audit. *Journal of economics & management strategy*, *20* (1), 3–41.

Lyon, T. P. & Montgomery, A. W. (2013). Tweetjacked: The impact of social media on corporate greenwash. *Journal of business ethics*, *118* (4), 747–757.

Lyon, T. P. & Montgomery, A. W. (2015). The means and end of greenwash. *Organization & Environment*, *28* (2), 223–249.

Macdonald, E. K. & Sharp, B. M. (2000). Brand awareness effects on consumer decision making for a common, repeat purchase product: A replication. *Journal of business research*, *48* (1), 5–15.

Majláth, M. (2016). How does greenwashing effect the firm, the industry and the society-the case of the vw emission scandal. *Proceedings of Fikusz 2016*, *111*.

Malhotra, N. K. & Dash, S. (2016). *Marketing research: An applied orientation*. Englewood Cliffs: Pearson.

Manju. (2012). Green marketing: New hopes and challenges. *Spectrum: A Journal of Multidisciplinary Research*, *1* (9), 40–47.

Mason, A. M. (2014). The impact of media frames and treatment responsibility within the situational crisis communication theory framework. *Corporate Reputation Review*, *17* (1), 78–90.

Matthes, J., Wirth, W., Schemer, C. & Pachoud, N. (2012). Tiptoe or tackle? the role of product placement prominence and program involvement for the mere exposure effect. *Journal of current issues & research in advertising*, *33* (2), 129–145.

Matthes, J. & Wonneberger, A. (2014). The skeptical green consumer revisited: Testing the relationship between green consumerism and skepticism toward advertising. *Journal of advertising*, *43* (2), 115–127.

McQuarrie, E. F. & Mick, D. G. (2003). The contribution of semiotic and rhetorical perspectives to the explanation of visual persuasion in advertising. *Persuasive imagery: A consumer response perspective*, 191–221.

McWilliams, A. & Siegel, D. (2001). Corporate social responsibility: A theory of the firm perspective. *Academy of management review*, *26* (1), 117–127.

McWilliams, A., Siegel, D. S. & Wright, P. M. (2006). Corporate social responsibility: Strategic implications. *Journal of management studies*, *43* (1), 1–18.

Meyers-Levy, J. & Malaviya, P. (1999). Consumers' processing of persuasive advertisements: An integrative framework of persuasion theories. *Journal of Marketing*, *63*, 45–60. Zugriff am 2022-07-12 auf http://www.jstor.org/stable/1252100

Miles, M. P. & Covin, J. G. (2000). Environmental marketing: A source of reputational, competitive, and financial advantage. *Journal of business ethics*, *23* (3), 299–311.

Milgrom, P. & Roberts, J. (1986). Price and advertising signals of product quality. *Journal of political economy*, *94* (4), 796–821.

Mills, E. (2009). A global review of insurance industry responses to climate change. *The Geneva Papers on Risk and Insurance-Issues and Practice*, *34* (3), 323–359.

Mio, C. & Fasan, M. (2012). Does corporate social performance yield any tangible financial benefit during a crisis? an event study of lehman brothers' bankruptcy. *Corporate reputation review*, *15* (4), 263–284.

Mishra, S. & Suar, D. (2010). Does corporate social responsibility influence firm performance of indian companies? *Journal of business ethics*, *95* (4), 571–601.

Mitchell, A. A. (1986). The effect of verbal and visual components of advertisements on brand attitudes and attitude toward the advertisement. *Journal of consumer research*, *13* (1), 12–24.

Mitchell, L. & Ramey, W. (2011). Look how green i am! an individual-level explanation for greenwashing. *Journal of Applied Business and Economics*, *12* (6), 40–45.

Mohammed, A. & Rashid, B. (2018). A conceptual model of corporate social responsibility dimensions, brand image, and customer satisfaction in malaysian hotel industry. *Kasetsart Journal of social sciences*, *39* (2), 358–364.

Mohr, L. A., Erolu, D. & Ellen, P. S. (1998). The development and testing of a measure of skepticism toward environmental claims in marketers' communications. *Journal of consumer affairs*, *32* (1), 30–55.

Monbiot, G. (2009). *It will take more than goodwill and greenwash to save the biosphere. the guardian newspaper, 6th january.* Zugriff auf https://www.theguardian.com/commentisfree/2009/jan/06/oil-business-climate-change (abgerufen am 10.04.2021)

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of consumer research*, *26* (4), 323–339.

Morris, S. A., Bartkus, B. R., Glassman, M. & Rhiel, G. S. (2013). Philanthropy and corporate reputation: An empirical investigation. *Corporate Reputation Review*, *16* (4), 285–299.

Musgrove, C. C. F., Choi, P. & Chris Cox, K. (2018). Consumer perceptions of green marketing claims: An examination of the relationships with type of claim and corporate credibility. *Services Marketing Quarterly*, *39* (4), 277–292.

Newburry, W. (2010). Reputation and supportive behavior: Moderating impacts of foreignness, industry and local exposure. *Corporate Reputation Review*, *12* (4), 388–405.

Newell, S. J., Goldsmith, R. E. & Banzhaf, E. J. (1998). The effect of misleading environmental claims on consumer perceptions of advertisements. *Journal of Marketing Theory and Practice*, *6* (2), 48–60.

Nolan, K. P., Gohlke, M., Gilmore, J. & Rosiello, R. (2013). Examining how corporations use online job ads to communicate employer brand image information. *Corporate Reputation Review*, *16* (4), 300–312.

Nyilasy, G., Gangadharbatla, H. & Paladino, A. (2014). Perceived greenwashing: The interactive effects of green advertising and corporate environmental performance on consumer reactions. *Journal of business ethics*, *125* (4), 693–707.

Obermiller, C., Spangenberg, E. & MacLachlan, D. L. (2005). Ad skepticism: The consequences of disbelief. *Journal of advertising*, *34* (3), 7–17.

Olsen, M. C., Slotegraaf, R. J. & Chandukala, S. R. (2014). Green claims and message frames: how green new products change brand attitude. *Journal of Marketing*, *78* (5), 119–137.

Orlitzky, M. & Swanson, D. L. (2012). Assessing stakeholder satisfaction: Toward a supplemental measure of corporate social performance as reputation. *Corporate Reputation Review*, *15* (2), 119–137.

Ottman, J. A., Stafford, E. R. & Hartman, C. L. (2006). Avoiding green marketing myopia: Ways to improve consumer appeal for environmentally preferable products. *Environment: science and policy for sustainable development*, *48* (5), 22–36.

Pappu, R., Quester, P. G. & Cooksey, R. W. (2005). Consumer-based brand equity: improving the measurement–empirical evidence. *Journal of Product & Brand Management*, *14* (3), 143–154.

Parguel, B., Benoît-Moreau, F. & Larceneux, F. (2011). How sustainability ratings might deter "greenwashing": A closer look at ethical corporate communication. *Journal of business ethics*, *102* (1), 15–28.

Parguel, B., Benoit-Moreau, F. & Russell, C. A. (2015). Can evoking nature in advertising mislead consumers? the power of "executional greenwashing". *International Journal of Advertising*, *34* (1), 107–134.

Patten, D. M. (2002). The relation between environmental performance and environmental disclosure: a research note. *Accounting, organizations*

and Society, *27* (8), 763–773.

Peterson, R. A. (1994). A meta-analysis of cronbach's coefficient alpha. *Journal of consumer research*, *21* (2), 381–391.

Pfau, M., Haigh, M. M., Sims, J. & Wigley, S. (2008). The influence of corporate social responsibility campaigns on public opinion. *Corporate Reputation Review*, *11* (2), 145–154.

Pled, V. & Iatridis, G. E. (2012). Corporate social responsibility reporting: evidence from environmentally sensitive indus-tries in the usa. *International Review of Accounting, Banking & Finance*, *4* (2), 61–99.

Polasek, W. (1997). Einführung in die stichprobentheorie und in die stichprobenverteilung. In *Schließende statistik* (S. 1–20). Springer.

Polonsky, M. J. (2011). Transformative green marketing: Impediments and opportunities. *Journal of business research*, *64* (12), 1311–1319.

Pomering, A. & Johnson, L. W. (2009). Advertising corporate social responsibility initiatives to communicate corporate image: Inhibiting scepticism to enhance persuasion. *Corporate Communications: An International Journal*, *4* (4), 420–439.

Ponzi, L. J., Fombrun, C. J. & Gardberg, N. A. (2011). Reptrak^{tm} pulse: Conceptualizing and validating a short-form measure of corporate reputation. *Corporate reputation review*, *14* (1), 15–35.

Porter, M. & van der Linde, C. (1995). Green and competitive: ending the stalemate. the dynamics of the eco-efficient econo-my: environmental regulation and competitive advantage. , *33*.

Prabhu, J. & Stewart, D. W. (2001). Signaling strategies in competitive interaction: Building reputations and hiding the truth. *Journal of Marketing Research*, *38* (1), 62–72.

Pritchard, M. & Wilson, T. (2018). Building corporate reputation through consumer responses to green new products. *Journal of Brand Management*, *25* (1), 38–52.

Raab, G., Unger, A. & Unger, F. (2009). *Methoden der marketing-forschung*. Wiesbaden: Gabler Verlag.

Raithel, S. & Schwaiger, M. (2015). The effects of corporate reputation perceptions of the general public on shareholder value. *Strategic management journal*, *36* (6), 945–956.

Rao, A. R., Qu, L. & Ruekert, R. W. (1999). Signaling unobservable product quality through a brand ally. *Journal of Marketing Research*, *36* (2), 258–268.

Richardson, A. J., Welker, M. & Hutchinson, I. R. (1999). Managing capital market reactions to corporate social resposibility. *International Journal of Management Reviews*, *1* (1), 17–43.

Roberts, P. W. & Dowling, G. R. (2002). Corporate reputation and sustained superior financial performance. *Strategic management journal*, *23* (12), 1077–1093.

Robertson, J. L. & Barling, J. (2013). Greening organizations through leaders' influence on employees' pro-environmental behaviors. *Journal of organizational behavior*, *34* (2), 176–194.

Royne, M. B., Martinez, J., Oakley, J. & Fox, A. K. (2012). The effectiveness of benefit type and price endings in green advertising. *Journal of Advertising*, *41* (4), 85–102.

Rubinstein, A. (2003). "economics and psychology"? the case of hyperbolic discounting. *International Economic Review*, *44* (4), 1207–1216.

Rumelt, R. P. (1995). Inertia and transformation. In *Resource-based and evolutionary theories of the firm: Towards a synthesis* (S. 101–132). Springer.

Ruxton, G. D. & Beauchamp, G. (2008). Time for some a priori thinking about post hoc testing. *Behavioral ecology*, *19* (3), 690–693.

Saeidi, S. P., Sofian, S., Saeidi, P., Saeidi, S. P. & Saaeidi, S. A. (2015). How does corporate social responsibility contribute to firm financial performance? the mediating role of competitive advantage, reputation, and customer satisfaction. *Journal of business research*, *68* (2), 341–350.

Salvioni, D. M. & Gennari, F. (2014). Corporate governance, sustainability and capital markets orientation. *International Journal of Management and Sustainability*, *3* (8), 469–483.

Sarstedt, M., Wilczynski, P. & Melewar, T. (2013). Measuring reputation in global markets–a comparison of reputation measures' convergent and criterion validities. *Journal of World Business*, *48* (3), 329–339.

Schmuck, D., Matthes, J. & Naderer, B. (2018). Misleading consumers with green advertising? an affect–reason–involvement account of greenwashing effects in environmental advertising. *Journal of Advertising*, *47* (2), 127–145.

Schmuck, D., Matthes, J., Naderer, B. & Beaufort, M. (2018). The effects of environmental brand attributes and nature imagery in green advertising. *Environmental Communication*, *12* (3), 414–429.

Schwartz, B. & Tilling, K. (2009). Iso-lating–corporate social responsibility in the organizational context: a dissenting interpretation of iso 26000. *Corporate Social Responsibility and Environmental Management*, *16* (5), 289–299.

Seele, P. & Gatti, L. (2017). Greenwashing revisited: In search of a typology and accusation-based definition incorporating legitimacy strategies. *Business Strategy and the Environment*, *26* (2), 239–252.

Segev, S., Fernandes, J. & Hong, C. (2016). Is your product really green? a content analysis to reassess green advertising. *Journal of Advertising*, *45* (1), 85–93.

Sethi, S. P. (1975). Dimensions of corporate social performance: An analytical framework. *California management review*, *17* (3), 58–64.

Shell. (2021). *Homepage.* Zugriff auf `https://www.shell.de` (abgerufen am 10.04.2021)

Skarmeas, D. & Leonidou, C. N. (2013). When consumers doubt, watch out! the role of csr skepticism. *Journal of business research*, *66* (10), 1831–1838.

Smith, K. T., Smith, M. & Wang, K. (2010). Does brand management of corporate reputation translate into higher market value? *Journal of Strategic Marketing*, *18* (3), 201–221.

Smith, V. & Langford, P. (2009). Evaluating the impact of corporate social responsibility programs on consumers. *Journal of Management & Organization*, *15* (1), 97–109.

Smith, V. L. & Font, X. (2014). Volunteer tourism, greenwashing and understanding responsible marketing using market signalling theory. *Journal of Sustainable Tourism*, *22* (6), 942–963.

Soleimani, A., Schneper, W. D. & Newburry, W. (2014). The impact of stakeholder power on corporate reputation: A cross-country corporate governance perspective. *Organization Science*, *25* (4), 991–1008.

Solomon, R. C. (1992). *Ethics and excellence: Cooperation and integrity in business*. Oxford, UK: Oxford University Press.

Spence, M. (1978). Job market signaling. In *Uncertainty in economics* (S. 281–306). Elsevier.

Sridhar, K. (2012). The relationship between the adoption of triple bottom line and enhanced corporate reputation and legitimacy. *Corporate Reputation Review*, *15* (2), 69–87.

Stephenson, E., Doukas, A. & Shaw, K. (2012). Greenwashing gas: Might a "transition fuel" label legitimize carbon-intensive natural gas development? *Energy Policy*, *46*, 452–459.

Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic management journal*, *17* (S2), 27–43.

Tai, F.-M., Chuang, S.-H. et al. (2014). Corporate social responsibility. *Ibusiness*, *6* (03), 117–130.

Taylor, C. R. (2015). A call for more research on "green" or environmental advertising. *International Journal of Advertising*, *34* (4), 573–575.

TerraChoice. (2010). The sins of greenwashing. *Ottawa, ON: TerraChoice Environmental Marketing Inc*.

Tomarken, A. J. & Serlin, R. C. (1986). Comparison of anova alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*, *99* (1), 90.

Torelli, R., Balluchi, F. & Lazzini, A. (2020). Greenwashing and environmental communication: Effects on stakeholders' perceptions. *Business Strategy and the Environment*, *29* (2), 407–421.

Torres, A., Bijmolt, T. H., Tribó, J. A. & Verhoef, P. (2012). Generating global brand equity through corporate social responsibility to key stakeholders. *International journal of research in marketing*, *29* (1), 13–24.

Treadway, D. C., Adams, G. L., Ranft, A. L. & Ferris, G. R. (2009). A meso-level conceptualization of ceo celebrity effectiveness. *The Leadership Quarterly*, *20* (4), 554–570.

Tucker, E. M., Rifon, N. J., Lee, E. M. & Reece, B. B. (2012). Consumer receptivity to green ads: A test of green claim types and the role of individual consumer characteristics for green ad response. *Journal of Advertising*, *41* (4), 9–23.

Ulrich, D. & Smallwood, N. (2007). Building a leadership brand. *Harvard Business Review*, *85* (7-8), 92–104.

Van den Bogaerd, M. & Aerts, W. (2014). Media reputation of a firm and

extent of trade credit supply. *Corporate Reputation Review*, *17* (1), 28–45.

Van Marrewijk, M. (2003). Concepts and definitions of csr and corporate sustainability: Between agency and communion. *Journal of business ethics*, *44* (2), 95–105.

Vázquez, R., Del Rio, A. B. & Iglesias, V. (2002). Consumer-based brand equity: development and validation of a measurement instrument. *Journal of Marketing management*, *18* (1-2), 27–48.

Veloutsou, C. & Moutinho, L. (2009). Brand relationships through brand reputation and brand tribalism. *Journal of Business Research*, *62* (3), 314–322.

Vlachos, P. A., Krepapa, A., Panagopoulos, N. G. & Tsamakos, A. (2013). Curvilinear effects of corporate social responsibility and benevolence on loyalty. *Corporate Reputation Review*, *16* (4), 248–262.

Vos, J. (2009). Actions speak louder than words: Greenwashing in corporate america. *Notre Dame JL Ethics & Pub. Pol'y*, *23*, 673–697.

Votaw, D. (1972). Genius becomes rare: A comment on the doctrine of social responsibility pt. i. *California management review*, *15* (2), 25–31.

Wade, J. B., Porac, J. F., Pollock, T. G. & Graffin, S. D. (2008). Star ceos: benefit or burden? *Organizational Dynamics*, *37* (2), 203–210.

Walker, K. (2010). A systematic review of the corporate reputation literature: Definition, measurement, and theory. *Corporate reputation review*, *12* (4), 357–387.

Walker, K. & Wan, F. (2012). The harm of symbolic actions and greenwashing: Corporate actions and communications on environmental performance and their financial implications. *Journal of business ethics*, *109* (2), 227–242.

Walsh, G. & Wiedmann, K.-P. (2004). A conceptualization of corporate reputation in germany: An evaluation and extension of the rq. *Corporate Reputation Review*, *6* (4), 304–312.

Walsh, J. P., Weber, K. & Margolis, J. D. (2003). Social issues and management: Our lost cause found. *Journal of management*, *29* (6), 859–881.

Wang, Q., Dou, J. & Jia, S. (2016). A meta-analytic review of corporate social responsibility and corporate financial performance: The moderating effect of contextual factors. *Business & society*, *55* (8), 1083–1121.

Waterhouse, B. C. (2017). The personal, the political and, the profitable: business and protest culture, 1960s-1980s. *Financial History* (121), 14.

Westphal, J. D. & Deephouse, D. L. (2011). Avoiding bad press: Interpersonal influence in relations between ceos and journalists and the consequences for press reporting about firms and their leadership. *Organization Science*, *22* (4), 1061–1086.

Wiedmann, K.-P. (2012). Ansatzpunkte zur messung der unternehmensreputation als grundlage einer erfolg versprechenden reputationsmanagementplanung–das reptrak-konzept als ausgangspunkt und skizzen zur relevanten weiterentwicklung. In *Corporate reputation management* (S. 57–101). Springer.

Wiedmann, K.-P., Fombrun, C. J. & van Riel, C. (2006). Ansatzpunkte zur messung der reputation von unternehmen. *Der Markt*, *45* (2), 98–109.

Wiedmann, K.-P., Hennigs, N., Schmidt, S. & Wuestefeld, T. (2013). Brand heritage and its impact on corporate reputation: Corporate roots as a vision for the future. *Corporate Reputation Review*, *16* (3), 187–205.

Wimbush, J. C., Shepard, J. M. & Markham, S. E. (1997). An empirical examination of the relationship between ethical climate and ethical behavior from multiple levels of analysis. *Journal of Business Ethics*, *16* (16), 1705–1716.

Wright, K. B. (2005). Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of computer-mediated communication*, *10* (3), JCMC1034.

Yoon, Y., Gürhan-Canli, Z. & Schwarz, N. (2006). The effect of corporate social responsibility (csr) activities on companies with bad reputations. *Journal of consumer psychology*, *16* (4), 377–390.