



Editor:

DOMINIK VAN AAKEN

Advisory Editorial Board:

FREDERIK AHLEMANN
JAN-PHILIPP AHRENS
THOMAS BAHLINGER
MARKUS BECKMANN
SULEIKA BORT
ROLF BRÜHL
KATRIN BURMEISTER-LAMP
CATHERINE CLEOPHAS
NILS CRASSEL
BENEDIKT DOWNAR
KERSTIN FEHRE
MATTHIAS FINK
DAVID FLORYSIK
GUNTHER FRIEDL
MARTIN FRIESL
FRANZ FUERST
WOLFGANG GÜTTEL
NINA KATRIN HANSEN
ANNE KATARINA HEIDER
CHRISTIAN HOFMANN
SVEN HÖRNER
STEPHAN KAISER
NADINE KAMMERLANDER
ALFRED KIESER
ALEKSANDRA KLEIN
NATALIA KLIEWER
DODO ZU KNYPHAUSEN-AUFSESS
SABINE T. KÖSZEGI
ARJAN KOZICA
CHRISTIAN KOZIOL
MARTIN KREEB
WERNER KUNZ
HANS-ULRICH KÜPPER
MICHAEL MEYER
JÜRGEN MÜHLBACHER
GORDON MÜLLER-SEITZ
J. PETER MURMANN
ANDREAS OSTERMAIER
BURKHARD PEDELL
ARTHUR POSCH
MARCEL PROKOPCZUK
TANJA RABL
SASCHA RAITHEL
NICOLE RATZINGER-SAKEL
ASTRID REICHEL
KATJA ROST
THOMAS RUSSACK
FLORIAN SAHLING
MARKO SARSTEDT
ANDREAS G. SCHERER
STEFAN SCHMID
UTE SCHMIEL
CHRISTIAN SCHMITZ
MARTIN SCHNEIDER
MARKUS SCHOLZ
LARS SCHWEIZER
DAVID SEIDL
THORSTEN SELLHORN
STEFAN SEURING
VIOLETTA SPLITTER
ANDREAS SUCHANEK
TILL TALAULICAR
ANN TANK
ANJA TUSCHKE
MATTHIAS UHL
CHRISTINE VALLASTER
PATRICK VELTE
CHRISTIAN VÖGTLIN
BARBARA E. WEISSENBERGER
ISABELL M. WELPE
HANNES WINNER
THOMAS WRONA
THOMAS ZWICK

JUNIOR MANAGEMENT SCIENCE

- Marie-Claire Joyeaux**, Work Less, Live More? The Impact of an Introduction of the Four-Day Working Week on Happiness in the Context of the Icelandic Four-Day Working Week Experiment 831
- Niklas Janßen**, Integrating Sustainability in Risk Management and Internal Control Systems: An Empirical Assessment of ESG Reporting of German DAX40 Firms 858
- Tobias Keserü**, A New Dimension of Transparency: ESG Disclosure and Its Effect on Shareholder Behavior 876
- Antonia Engel**, ESG Regulation Across the Globe: Does ESG Regulation Pay Off? 904
- Finn Matthes Gooßen**, The Impact of Female Board Members on ESG Performance: An Empirical Analysis 940
- Louisa Felicitas Bläßer**, Implicit Measurement of the Moral Self-Image Using the Go/No-Go Association Task (GNAT) - An Empirical Investigation of the Convergent Validity Between Explicit and Implicit Measures 966
- Laura Wiredu**, Who Bears the Costs of the UK Soft Drink Tax? An Empirical Study of Medium-Term Effects 985
- Thiemo Janßen**, From Pictures to Perceptions: Exploring the Strategic Use of Visuals in CSR Reports and the Impact of Regulatory Mandates 1009
- Vanessa Jeske**, Determinants of Corporate Bond Mutual Fund Flows 1028
- Felix Yumuşak**, The Influence of Leadership Style on the Acceptance of Generative AI in the Workplace - The Role of Organizational Commitment, Job Insecurity and Interaction Frequency 1053

Published by Junior Management Science e.V.

This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.



Work Less, Live More? The Impact of an Introduction of the Four-Day Working Week on Happiness in the Context of the Icelandic Four-Day Working Week Experiment

Marie-Claire Joyeaux

University of St Andrews

Abstract

In our current age, the quest for a better work-life balance is becoming paramount to increasing numbers of people. The concept of the Four-Day Working Week (FDWW) therefore emerges as a potential solution, promising to revolutionise our traditional understanding of worktime and well-being. By focusing on the world's largest FDWW trial at the time, Iceland's FDWW experiment from 2015 to 2019, and using data from the European Social Survey, the FDWW's impact on happiness is critically examined. Ordinary Least Squares regressions are employed to analyse the association of working hours with happiness amongst Icelandic workers and to conduct an evidence-based policy evaluation. Contrary to widespread expectations and existing media narratives, the findings reveal no significant impact of working hours on employees' happiness scores. This outcome challenges the conventional wisdom that less work leads directly to more happiness. The findings thereby contribute significantly to debates on the future of work, suggesting that the FDWW should be viewed with caution until more conclusive evidence is available. This absence of definitive proof calls into question the notion of the FDWW as a universally effective solution to the economic and social challenges faced by contemporary society.

Keywords: employee well-being; four-day working week; future of work; workplace happiness; work time reduction

1. Introduction

Work less, live more? Today's rapidly evolving work environment is an era where the boundaries between work and life blur, focused on the relentless pursuit of progress and productivity and leading society to overlook the fundamental question of whether more work truly leads to a more fulfilling and happier life. As the new generation Z (born 1997-2012), who likes to work autonomously in a flexible workplace, slowly enters the labour market and challenges this long working hours culture of modern societies and liberal

welfare states (Bulut & Maraba, 2021), the *Four-Day Working Week* (FDWW) emerges as a contemporary concept that places the notion of less work and more life outside of work in the foreground. It serves not only as a policy experiment, but as a beacon of hope that has the potential to bridge gaps and provide a solution to pressing matters of individuals, organisations, and politics alike (T. T. Campbell, 2023), such as well-being, labour shortages, or environmental concerns.

The idea of reducing working hours has its roots in the labour movements of the early 20th century, gaining significant traction in Western Europe during the post-World War II period (T. T. Campbell, 2023; Mahoney, 1978). Notably, Scandinavian countries are early adopters of progressive labour policies, laying the groundwork for contemporary discussions around the potential of the FDWW as a novel approach that challenges traditional work models by reducing working hours while paying workers a full wage (Chung, 2022). Nonetheless, despite receiving positive media coverage, the scientific evidence supporting the benefits of the

My sincere thanks go to the Department of Finance & Economics at the University of St Andrews for fostering an intellectually open and supportive environment, which enabled me to pursue this work at a high level of quality with independence and purpose. Above all, I am deeply grateful to my family for their unwavering support, encouragement, and love throughout this journey. Special thanks to my wonderful St Andrews friends, whose kindness and companionship sustained me in more ways than words can say

FDWW remains inconclusive. Modern happiness economics argues that there is a positive aggregate effect of work that increases life satisfaction (Frey et al., 2007), while neoclassical economic theory points to the disutility effect of labour at the margin (Rätzel, 2012), highlighting inconsistent results across various studies (Shao, 2022). Therefore, the current literature fails to identify consistent effects of reduced working hours on happiness, as the state of evidence is not robust, often affected by methodological limitations and a lack of comprehensive data (T. T. Campbell, 2023). Additionally, since the concept of a reduced working week is not merely a logistical change, but represents a fundamental shift in how society views the balance between work and leisure, it becomes of importance to investigate both the underlying relationship between working hours and happiness and, consequently, the effect of the FDWW itself. The interest in answering these associations is motivated by the current state of evidence surrounding working hours and happiness, which suggests that while the effects of working hours on overall life satisfaction (Shao, 2022) and different outcomes like income (Clark et al., 2008) or productivity (Golden, 2012) are well-studied, the direct relationship between (reduced) working hours and specifically happiness remains underexplored (Golden & Wiens-Tuers, 2006). It is proven that there is some association, however, existing studies offer only conflicting results or are limited in scope (for example, see Booth and van Ours, 2008; Meier and Stutzer, 2004; Pouwels et al., 2008).

Accordingly, it becomes of interest to study the association of the worktime - happiness nexus independent of the true question of causality, to determine whether a reimagined work schedule can indeed reshape the dynamics of workplace happiness. The world's largest FDWW trial at the time, conducted nationwide in Iceland between 2015 and 2019 and involving 1.3% of Iceland's total workforce (Haraldsson & Kellam, 2021), provides a unique empirical setting to carry out an evidence-based policy evaluation that addresses the research question of whether the FDWW has the potential to provide greater happiness. More precisely, in the context of the Icelandic FDWW trial, the aim of this thesis is to provide a comprehensive analysis that contextualises the trial's outcomes within the broader theoretical frameworks of working hours and happiness. This contribution to ongoing discussions about the role of working hours in relation to happiness and insights into the applicability of the FDWW could not only guide future academic discourse, but also inform policy-making on a global scale. Therefore, this research adds to this field by considering empirical evidence from the Icelandic FDWW trial, exploring the nuanced relationship between working hours and happiness, and highlighting the potential and limitations of a FDWW as a policy measure to enhance well-being.

A weighted, clustered, and stratified analysis, employing Ordinary Least Squares (OLS) regressions to detect effects of working hours on happiness is conducted, based on the European Social Survey (ESS) dataset on Iceland (ESS, 2024b). The findings of this thesis from the analysis in the

Icelandic context provide valuable insights, highlighting a paradox contrary to some expectations, as working hours do not show a straightforward or significant impact on overall employee happiness. This counterintuitive outcome challenges prevailing assumptions within both academic circles and practical policymaking, namely that less work automatically leads to more happiness, suggesting a more complex interaction between working hours and happiness. The outcomes of this thesis therefore raise pivotal questions about (i) standard economic models, as the results indicate that happiness could be increased by more than just leisure time, and especially about (ii) the applicability of the FDWW concept, as no evidence of its success can be presented. In fact, there is no evidence of potential benefits of the FDWW to begin with, challenging the positive voices from the media and suggesting that this concept is rather a privilege of modern Western states than a future-proof concept for all.

To facilitate the analysis, the thesis is divided into four main sections that will explore the above-mentioned research question. The first section reviews the relevant literature around happiness economics and working hours, thereby setting the stage for a deeper examination of the FDWW trial by contextualising the thesis' topic within broader economic and social debates. Then, the methodology chapter details the dataset and analytical strategies employed, linking theory to the analytical framework used for analysis. Next, the results section explores the core findings from the analysis, considering several model specifications and robustness checks to ensure reliability and validity of the findings. Finally, the discussion section engages in a thorough analysis of the results, examining limitations of this work, integrating findings with the existing body of research, and considering future prospects to draw conclusions about the topic at hand.

2. Literature Review

This literature review sets the stage for understanding the contemporary discourse surrounding the FDWW experiment in Iceland and its potential impact on happiness by delving into two intertwined domains. First, the multidimensional aspects of happiness within happiness economics are analysed, and second, the historical shifts in work dynamics are traced, in order to establish hypotheses, the research question, and research objective.

2.1. Happiness Economics

The field of happiness economics is expanding globally, exploring the relationship between happiness and various economic factors. The following subchapters offer an exhaustive review of the literature, covering definitions, methods and critique, applications, and theory.

2.1.1. The Concept of Happiness and Happiness Economics

A comprehensive understanding of *subjective well-being* (SWB), a multidimensional concept that encompasses distinct ways of how a person's well-being is perceived (A.

Campbell, 1976), is provided by considering two distinct components, namely cognitive well-being and affective well-being (Diener, 1984; Luhmann, Hawkey, et al., 2012). Whereas cognitive well-being refers to the cognitive assessment of one's own life satisfaction (Diener et al., 2010), affective well-being involves the affective judgement of positive and negative moods and emotions (Luhmann, Hofmann, et al., 2012), with both elements critically influencing overall life quality evaluations, drawing on standards of a good life (cognitive) and hedonic affect levels (Veenhoven, 2008).

Under the umbrella term of SWB, two different concepts are united, *happiness* and *satisfaction*, which scholars often confuse and use interchangeably despite being distinct, as life satisfaction concerns cognition, while happiness concerns affect (Golden & Okulicz-Kozaryn, 2015). While the pursuit of happiness has been a central theme throughout all of human history, it has only become a focus of diverse scientific studies since the mid-20th century (McMahon, 2008), yielding varied definitions influenced by disciplines, historical, and cultural contexts (Frey, 2018). For this thesis, it is important to take into account that happiness, which pertains more to affect and short-term emotional states, is particularly sensitive to changes in daily routines like those introduced by the FDWW, as emotional responses are often more sensitive to adjustments in work-life balance than long-term cognitive evaluations like life satisfaction.

The general understanding of the term happiness is considered to be individualistic, in its broadest sense defined by Michael Argyle and Jill Crossland (1987) as an individual having a positive inner feeling towards particular aspects in life and drawing on cognitive and judgmental thoughts to approach both positive and negative emotions, ultimately constituting an individual's happiness (Andrews & Withey, 1976). The overlapping definitions of life satisfaction and happiness do not only show that the concepts are closely related to each other, but moreover highlight the reason for the lack of effort in the literature to distinguish between the two terms (MacKerron, 2012). The distinction, however, is critical for this thesis where happiness is chosen to assess SWB, as it provides a more immediate and affective measure of well-being (Kahneman & Deaton, 2010). For while life satisfaction is a valuable measure in many contexts, it may not accurately reflect the direct impact of workplace interventions like reduced working hours, which are more likely to have a noticeable effect on daily happiness, that is, people's affect (Krueger et al., 2005), rather than on a comprehensive assessment of life satisfaction.

The burgeoning research field of happiness economics has consequently emerged out of applying these concepts in an economic context. By examining the psychological underpinnings of well-being alongside economic indicators such as income, employment, and work hours, this interdisciplinary approach clarifies how economic conditions impact SWB, offering a nuanced understanding of the relationship of happiness in the intersection between economics and psychology (Nikolova & Graham, 2020). Furthermore, the idea of happiness economics is built upon the broader understanding of

utility and welfare, including interconnected utility functions and emphasising the significance of non-pecuniary life events in shaping individual's SWB (Graham, 2005), recognising happiness and well-being as crucial outcomes. The first modern economist to re-visit the concept of happiness, Richard Easterlin (2005), emphasised the significance of happiness economics, critiquing traditional economic ("more is better") and psychological theories (set point theory) for their inability to adequately explain SWB, as they fall short in accounting for the enduring impacts of non-pecuniary life events. In particular, traditional neo-classical economics relies primarily on the preference satisfaction model of well-being, linking happiness to the ability to fulfil one's desires, whereas the emerging field of happiness economics prioritises both the evaluative (cognitive) and hedonic aspect, giving precedence to individuals' assessment of their own well-being (MacKerron, 2012).

This subjective evaluation of one's happiness is self-reported, hence prompting concerns about the validity of its measure and explanatory power in scientific research (Norrish & Vella-Brodrick, 2008). Numerous issues arise that may skew results: cultural disparities might lead to different interpretations of what constitutes happiness across societies; social desirability bias can cause respondents to answer in ways they believe are socially acceptable rather than truthful; memory bias may affect how past events are recalled, influencing reported happiness levels; and mood-state dependence implies that a person's current emotional state can disproportionately influence their overall happiness assessment (Diener et al., 2018; MacKerron, 2012). These factors challenge the reliability of happiness measures, however, subjective measures of happiness appear to be reasonable, robust, and reliable proxies for the traditional economic concept of utility, and are hence valid measures used by economists (Di Tella & MacCulloch, 2006; Dolan & White, 2007; Frey & Stutzer, 2002; Krueger, 2009; Stiglitz et al., 2009) and in this thesis.

2.1.2. Determinants of Happiness

An individual's happiness is determined by a number of factors which happiness research seeks to identify, isolate, and measure (Frey, 2018). Economic literature broadly groups these factors into three different categories, namely (i) economic factors, (ii) individual factors, and (iii) social factors.

To begin with, the extensively researched relationship between income and happiness (Clark et al., 2008) reveals a complex dynamic. While higher average income generally correlates with greater SWB both at individual (Killingsworth, 2021) and national (Frey, 2018) level, this association is found to only hold up to an income threshold of 75,000 US Dollar per year, after which non-financial factors become of relevance (Kahneman & Deaton, 2010). The significance often lies more with relative income rather than absolute income (Easterlin, 1974) and findings indicate a positive relationship that may stem from reverse causation (Diener et al., 2002; Graham et al., 2004), however, with

diminishing marginal returns over time (Frey, 2018). As marginal utility of income decreases while income rises, processes like social comparison and hedonic adaptation are put forward to explain why more wealth reduces an individual's life satisfaction (Stutzer, 2004). This leads to the Easterlin Paradox, a term used to explain the phenomenon that over time there is no corresponding effect of rising income on the long-term growth rate of happiness (Easterlin, 1974; Easterlin & O'Connor, 2020).

Furthermore, concerning individual factors, age is consistently shown to have a negative correlation with SWB while age squared has a positive association with SWB, with an indicated U-shaped relationship suggesting years of mid-life crisis, that is, higher levels of SWB in younger and older age groups and lowest SWB in the middle ages of 32 to 50 (Blanchflower & Oswald, 2004; Ferrer-i-Carbonell & Gowdy, 2007). Additionally, gender effects on happiness are mixed: whereas some studies report higher levels of happiness for women (Alesina et al., 2004), others find no significant gender differences (Louis & Zhao, 2002), indicating that other factors beyond gender itself may be more influential (Dolan et al., 2008). Besides, scholars suggest a certain influence of ethnicity on happiness, finding whites in the United States to report higher SWB than African Americans (Thoits & Hewitt, 2001), however, a comparison across diverse groups remains complex due to broad ethnic categorisations (Luttmer, 2005).

Last but not least, social factors play a crucial role in determining happiness. In particular, both an individual's subjective (Howell & Howell, 2008) and objective socioeconomic status (such as access to essential services like healthcare) can significantly influence happiness by affecting an individual's physical and mental well-being (Kezer & Cemalcar, 2020). Literature is unclear about the influence of education on SWB. On one hand, education is analysed to open up more opportunities in life and thus contributing to greater happiness (Blanchflower & Oswald, 2004), but on the other hand also no significant relationship (Flouri, 2004) or even a negative relationship (Clark, 2003) is reported. As the variable of education is mostly positively correlated with income and health, a strong positive education coefficient may pick up unobserved effects that are not controlled for (Dolan et al., 2008). Furthermore, general health is one of the most important contributors to happiness, as studies prove strong positive relationships between SWB and both physical and psychological health (Dolan et al., 2008), that is, (poor) health strongly predicting (un)happiness (Liu et al., 2015). Moreover, marriage is examined to be highly positive with SWB (Helliwell, 2003), as married people or those in stable partnerships feel less lonely, with these relationships acting as a buffer against stressors of work life (Frey, 2018). Finally, research is very consistent in suggesting that individuals having faith and trust in a higher power are more adept at dealing with life's challenges (Frey, 2018), and that regular church attendance is associated with higher SWB (Clark & Lelkes, 2006).

2.1.3. Working Hours in the Intersection of Traditional and Happiness Economics

While a considerable portion of research on SWB examines the impact of one of the above three categories, very little research has turned towards the time spent working or in a workplace (Golden & Okulicz-Kozaryn, 2015). A rich body of research is found within occupational psychology, with a wealth of evidence that employees who work long or extended working hours encounter more burnout, stress, anxiety, sleep shortage, and increased illness (Bannai & Tamakoshi, 2014; Chung, 2022; Kleiner & Pavalko, 2010; Wong et al., 2019); research in happiness economics, however, has given little consideration to the impact of working hours on SWB (Golden & Wiens-Tuers, 2006) and - if at all - focused on the relationship between working hours and job satisfaction, a subarea of general satisfaction (Shao, 2022).

This blind eye towards the worktime - happiness nexus may be attributed to the mismatch of the standard neoclassical theory of individual labour supply and emerging research in the field of happiness economics. Empirical findings within happiness economics indicate that unemployment severely impacts SWB negatively (Frey & Stutzer, 2002; Helliwell, 2003), highlighting a positive aggregate effect of work, as being employed increases life satisfaction (Frey et al., 2007) through psychological and sociological factors such as self-confidence (Frey, 2018). The confirmed non-pecuniary cost of unemployment associated with the hardship of job loss (Young, 2012) suggests that a utility reduction of an individual caused by unemployment cannot be wholly attributed to the loss of income (Rätzel, 2012), that is, happiness does not consistently rise with income, especially in higher income groups (happiness paradox), but is rather contingent upon other factors such as health and family life experience (Golden & Wiens-Tuers, 2006). On the other hand, standard neoclassical economic theory of individual labour supply follows the notion that more working hours increase an individual's income, hence allowing for the possibility to pay for health and thereby outweighing any adverse effects of labour disutility (Golden & Wiens-Tuers, 2006). This disutility effect as described by conventional economics posits labour disutility at the margin and diminishing marginal returns, namely that disutility is caused by an additional hour of work (Rätzel, 2012) which, at first glance, seems to contradict the positive aggregate effect of work as found in happiness economics.

Thus, recent research has attempted to reconcile both theories by distinguishing between total and marginal labour utility (Rätzel, 2012), resulting in mixed findings only and thus calling for more in-depth research that has yet to be undertaken (Shao, 2022). So far, there is evidence to suggest that there is some relationship between working hours and happiness, this relationship, however, is not consistent. Some find the well-being of individuals to not decrease with longer working hours even though job satisfaction decreases (Gray et al., 2004). Scholars also discover no significant relationship (Booth & van Ours, 2008; Willson & Dickerson, 2010), others observe the relationship to be negative

(Luttmer, 2005; Pouwels et al., 2008), or non-linear (Frey et al., 2007; Meier & Stutzer, 2004; Muffels & Kemperman, 2011; Rätzel, 2012), stressing an inverse U-shaped relation between happiness and hours worked, in other words, implying that well-being increases with working hours up to a certain threshold after which it begins to decline with excessive working hours (Meier & Stutzer, 2004). As per utility theory (Viner, 1925), workers maximise their utility by seeking a balance between leisure time and income, achieving peak satisfaction at an optimal number of working hours, with dissatisfaction occurring when deviating from this balance due to insufficient income or leisure (Dong et al., 2023).

This background about the impact of working hours on happiness leads to the formation of the first hypothesis:

H₁: There is an inverse U-shaped relationship between the number of working hours and happiness.

2.2. Work Dynamics Over Time

The concept of happiness has taken hold in the workplace in the 21st century as organisations start prioritising human capital to sustain their presence in the workforce and enhance efficiency (Arslan & Bektaş, 2021), inducing significant changes in working time regulations. Therefore, the following subchapters examine changes in working time patterns over time, with a focus on the definition and phenomenon of the concept of the FDWW.

2.2.1. Historical Context of Working Hours

Adam Smith's early theory of "Economic Man", advocating for indefinite extension of employees' working hours to maximise profit (Liu et al., 2018), is long foregone, with the new post-millennial generation demanding more freedom, flexibility, and diverse ways of working and collaborating (Bulut & Maraba, 2021). Therefore, understanding historical perspectives of Western Europe is essential for comprehending contemporary trends in working hours.

First, the beginning of the period of 1900-1990 in Western Europe is characterised by fluctuating working hours with unstable wages, sensitive to economic disturbance caused by the two World Wars (Liu et al., 2018). Following World War II, increased female participation in part-time roles in the labour market prompts a decline in overall working hours (Stier & Lewin-Epstein, 2003), which, next to intensified economic competition, spurs interest in more flexible and diverse working arrangements, predominantly flextime (Mahoney, 1978), part-time work (Liu et al., 2018), and a compressed work week, that is, keeping working hours constant, but compressing the 40h work week into four work days (Allen & Hawes, 1979; Mahoney et al., 1975).

Subsequently, intensified global competition and cooperation during the years 1990-1999 due to changes in political and economic powers lead to more pressure of overtime work (Liu et al., 2018) and shift work, causing tensions between organisations and individuals and thus an increased interest in work-life balance (Anttila et al., 2021). Nonetheless, research on reduced working time arrangements stalls

(T. T. Campbell, 2023) due to resistance to adapt to unfamiliar schemes (Smith, 1986), except for studies on compressed workweeks, reviewed to be positively correlated with job satisfaction, productivity, and employee absenteeism (Moores, 1990).

Thereafter, during the years 2000-2012, research on working time regulations spikes, as the global financial crisis and rising fuel and energy prices highlight benefits of reduced commuting and operating costs (T. T. Campbell, 2023), as well as environmental pressures (Kallis et al., 2013). Additionally, academic research connects positive employee experience and increased productivity (Facer & Wadsworth, 2008; Golden, 2012), motivating the interest in transitioning towards a shorter workweek.

Last but not least, as the work-life conflict becomes of greater importance after 2010, there is extensive research on long working hours and consequential psychological distress, with emphasis on economic gains of long working hours being significantly outweighed by the social costs they entail (Liu et al., 2018). A second contemporary wave of interest in shorter working weeks that has a particular focus on the concept of the FDWW resurges around 2019, however, scholarly publications on the topic remain limited, as the interest is primarily driven by media, trade unions, political parties, think tanks, and advocacy groups, rather than by academia (T. T. Campbell, 2023).

This historical context highlights the gap in rigorous empirical research despite over a century of widespread public and political interest, underscoring the relevance of this thesis in contributing to a well-informed academic debate on the topic. Especially since the COVID-19 pandemic, future trends in working hours have been shaped significantly, not only by accelerating shifts towards telework, but above all by stimulating dialogue on more flexible and diversified employment patterns as a means of tackling unemployment, demographic transition, and climate change (Anttila et al., 2021).

2.2.2. The Concept of the Four-Day Working Week

The traditional five-day eight-hour working week has long been a global standard for many companies; however, the idea of a FDWW has gained popularity amongst the younger generations as a potential solution to various workplace challenges (T. T. Campbell, 2023). The concept of *work time reduction* (WTR) stands for a reduction in the total hours of paid work (Gunderson, 2018) and is an umbrella term, encompassing various working time regulations such as part-time work, compressed work weeks, and the specific approach as proposed by the FDWW. Unlike other WTR arrangements, the FDWW entails a reduction of working hours which will be the new equivalent to full-time working hours, that is, a full working week of around 36-40 hours should move to a new standard of 30-32 hours without cutting payments of workers (Chung, 2022).

Recent years have seen increasing interest in FDWW trials across the globe, with companies like Microsoft Japan (2019) conducting a FDWW experiment (Gatlin-Keener & Lunsford, 2019), as well as entire countries running FDWW

trials. This innovative approach to work, however, has not yet become a common practice in mainstream corporate environments, including liberal welfare states, where prevailing norms still emphasise the need to work longer hours as a demonstration of dedication, ambition, productivity, performance, and commitment (Deci & Ryan, 2009; Laurence et al., 2016). Consequently, many employees either extend their work hours or compensate for absent colleagues (Barrick & Mount, 1991; Hayes et al., 2022), despite evidence suggesting that long working hours reverse organisational performance and are negatively related to economic development (Liu et al., 2019).

Therefore, a surge in scientific research on the impact of WTR has been observable in recent years, offering valuable insights for policymakers, employers, and employees on how WTR drives economic, organisational, and personal change. As such, the positive link of WTR to productivity (Kunn-Nelen et al., 2013; Pencavel, 2015), as well as to climate change mitigation (Kallis et al., 2013; Nässén & Larsson, 2015; Pullinger, 2014) has been studied extensively, as there are reasons to believe that reducing working hours can improve quality of life and promote sustainable living by shifting the focus from living to work with carbon-intensive consumption to relationship, hobbies, and local engagement (Coote et al., 2010; Schor, 2015). Next to this, the negative impact of long working hours on family balance (Crouter et al., 2001; Johnson et al., 2013; Strazdins et al., 2017) and well-being (Bannai & Tamakoshi, 2014; Caruso et al., 2006; Sonnentag, 2012; Wong et al., 2019) has been researched in detail. In particular, research consistently indicates that long working hours have notable detrimental effects on employees' well-being by having a direct negative impact on their physical and mental health (Chung, 2022).

In spite of these findings, it becomes apparent that much of this research tends to generalise findings to the FDWW scenario without adequately considering potential differences in effects compared to other forms of WTR (T. T. Campbell, 2023), inducing the need of more studies that focus on the FDWW concept itself.

Based on above analysed dynamics of working hours in the context of the FDWW, a second hypothesis is proposed:

H₂: The association between working hours and happiness changes before, during, and after the FDWW trial in Iceland.

2.3. The Four-Day Working Week in the Context of the Icelandic Experiment

Quoted to be an “overwhelming success” by all newspapers, the world's largest ever FDWW trial at the time was conducted in Iceland and has gained traction for being a paragon in promoting the idea of a FDWW nationwide in an organisational context (Stone, 2021).

Between 2015 and 2019, trials took place in Iceland, involving a reduction of the 40-hour work week to a 35-hour week without cutting payments of employees. In particular, workplaces were selected based on fulfilling the criteria

of having (i) 20 or more full-time employees, (ii) at least 30% of the employees in BRSB member companies, that is, in companies that are part of Iceland's largest federation of public sector unions, (iii) jobs performed at the workplace that are quite similar, and (iv) a majority of employees at a 70% - 100% employment rate (Félagsmálaráðuneytið, 2019). From an initial 17 applications four workplaces were selected, with another four as a control group. The scope of the participating workplaces was then expanded until more than 100 different workplaces were involved, including offices, kindergartens, social service providers, and hospitals, thereby encompassing more than 2,500 staff members equating to about 1.3% of Iceland's total workforce (Haraldsson & Kellam, 2021).

That a FDWW without a salary drop can indeed have a powerful positive effect was confirmed by the trial through empirical evaluations of employee feedback and assessments of qualitative surveys, with researchers claiming that productivity stayed the same or improved in most of the workplaces and employees reporting that the shortage of working hours increased their work-life balance (Haraldsson & Kellam, 2021). However, despite the apparent success of the trial, the research design does not allow for a definitive treatment effect, as differences in outcomes cannot be attributed directly to the treatment of reduced working hours. Whereas the data show that working hours have been reduced, there is an over-reliance on opinion data and further, academic research has recognised that benefits of a FDWW are unreasonably positive, concluding that relationships are often oversimplified and that establishing a causal link requires further investigation (T. T. Campbell, 2023). This limitation underscores that while the experiment serves as a paragon for introducing changes in working hours, it principally provides an opportunity to study correlation rather than causation.

Nonetheless, the impact of Iceland's FDWW trial is far-reaching, as at the time of the report's publication in 2021, 86% of Iceland's entire workforce has already moved to a shorter working week or is negotiating shorter working hours with the help of Icelandic's trade unions and their confederations (Haraldsson & Kellam, 2021), potentially setting a desirable precedent for other countries and organisations considering the implementation of a FDWW in the future.

2.4. Research Question and Research Objective

The recognition that future of work is shaped by values and expectations of the new generation makes it necessary to explore how changes in working time regulations are perceived and desired by the workforce. Given that working hours have only been sporadically analysed in relation to happiness - the literature focuses almost exclusively on the income - happiness nexus (Yamashita et al., 2016), and that the relationship of causality that can run both ways remains of debate, the only remaining question of interest, independent of the true question of causality, brings the direction, form, and strength of the relationship to the fore, in order to answer whether there is an optimal number of working

hours that induce the highest happiness amongst employees (Kameråde et al., 2019), and if so, at what point.

Therefore, from the hypotheses constructed based on above literature review, the research question to examine is whether the FDWW has the potential to provide a solution to greater happiness. The research objective is twofold, aiming (i) at contributing to the greater context of discussion about the relationship of working hours and happiness, and (ii) at conducting an evidence-based policy evaluation by providing insights into the potential impact of the FDWW on employee well-being and analysing the association between working hours and happiness before, during, and after the FDWW trial in Iceland. In addition, an outlook on how a shift to a FDWW can contribute to offering an opportunity to solve issues centring around contemporary problems faced by society and individuals is explored. All in all, by critically analysing the worktime - happiness nexus in the context of Iceland and consequently identifying gaps in the state of research around the shape of the working hours - happiness relationship and concepts of future of work, this thesis aims at highlighting the importance of the Icelandic experiment as a paragon in helping to answer the overarching economic question of labour (dis)utility and its relation to hours worked. Finally, this thesis seeks to draw attention to the need to consider other factors beyond working hours in promoting happiness in the workplace, and to inform discussions about the efficacy of FDWW trials.

3. Data and Methodology

The present thesis employs an OLS regression analysis in order to rigorously examine and identify the association between reduced working hours and happiness in Iceland's FDWW experiment. Given the theoretical background established, the following subchapters detail the data source for this analysis and the utility theory-based analytical framework employed to examine this relationship.

3.1. The European Social Survey Dataset and (Transformation of) Variables

The analysis of the worktime - happiness nexus conducted in this thesis is based on data from the ESS, a biennial academically cross-national survey that has set a particular focus on collecting responses on social indicators across Europe since 2001 (ESS, 2024a). The ESS employs a rigorous, robust sampling strategy to ensure accurate and representative data for all individuals aged 15 and over within each participating country. Therefore, the sampling designs includes both single-stage sampling in urban areas and multi-stage sampling in rural areas to address geographic and demographic complexities. Additionally, preferred sampling frames include population registers, allowing for equal-probability sampling, supplemented by lists of addresses or area sampling when registers are unavailable. Furthermore, data collection in the ESS is conducted through face-to-face computer assisted personal interviewing interviews, ensuring

high-quality data acquisition (ESS, 2024c). Consequently, these methodologies collectively ensure consistent data collection, affirming the ESS as a reliable source for this research on the impact of working hours on happiness.

The thesis uses repeated pooled cross-sectional data from Iceland, sourced from the ESS website (ESS, 2024b) in the available years 2004, 2012, 2016, 2018, and 2020, capturing periods before, during, and after Iceland's FDWW trial to analyse changes in public perceptions and behaviours across these periods. Therefore, the analysis leverages a happiness scale from 0 to 10 as the dependent variable, based on the happiness question that reads: *"Taking all things together, how happy would you say you are?"*, with employees' total hours normally worked per week in their main job as the independent variable. Additional control variables that are relevant socio-demographic and socio-economic characteristics as established by literature include income, gender, age, marital status, religion, health, education, and belonging to an ethnic minority, which are documented alongside survey questions and indicator descriptions in Table 1.

Moreover, certain variables are transformed and adjusted: First, household income is equivalised according to the OECD Modified Scale which considers household size and composition - assigning a value of 1 to the household head, of 0.5 to each additional adult member, and of 0.3 to each child (OECD, 2009). To accurately apply this scale, midpoints from each income band based on the income categories shown to respondents are used (Kuhn, 2019), approximating household income in Euros (see Appendix A, p. 74). This procedure allows to create an income variable that considers the specific needs of each household member and takes into account economies of scale in consumption, enabling a fair cross-sample comparison of living standards (OECD, 2009). Further, to address skewed income distributions and the diminishing marginal utility of income as levels rise (Easterlin, 1974; Stevenson & Wolfers, 2013), the equivalised household income is log-transformed to achieve a normal distribution of the data, essential for statistical accuracy (Frey & Stutzer, 2002). The final analysis ultimately employs only the log-transformed income due to its high intercorrelation found (see Table 3). Second, for the variable weekly working hours, answers above 80 hours per week are considered implausible and set as missing due to Icelandic law limiting employees to working 162.5 hours per month, that is, 40.6 hours per week (VR stéttarfélag, n.d.), and even considering overtime and self-employed workers not bound by law, responses above 80 hours per week appear unrealistic. A further analysis of the distribution and effects of working hours shows few answers exceeding this threshold, and including or excluding these observations does not significantly alter the results, confirming the decision taken. In addition, the dataset includes two distinct variables for working hours, namely total contracted hours excluding overtime and total hours including overtime. A comparison of models and post-estimation marginal effects for each variable helps in determining which variable better represents the data, with the overtime-inclusive variable ultimately being

Table 1: Variables Used for Regression Analysis, Their Scales, and Definition

DESCRIPTION OF VARIABLES		
Variables	Scale	Survey Question / Definition
Dependent Variable		
Happiness (happy)	0-10 0 = Extremely Unhappy; 10 = Extremely Happy	Taking all things together, how happy would you say you are?
Independent Variables		
Weekly Working Hours incl. Overtime (work_hours_ot)	[0 ; 80]	Regardless of your basic or contracted hours, how many hours do/did you normally work a week (in your main job), including any paid or unpaid overtime?
Squared Weekly Working Hours incl. Overtime (work_hours_ot_squared)	[0 ; 6,400]	Squared total contracted hours per week in main job, overtime included
Weekly Working Hours incl. Overtime in Categories (hours_category)	[0 ; 5] 0 = Hours 0 ; 1 = Hours 1-30 ; 2 = Hours 31-40 ; 3 = Hours 41-50 ; 4 = Hours 51 - 60 ; 5 = Hours 60+	Weekly working hours per week in main job, overtime included, split in categories
Age (age)	[21 ; 67]	Age of respondent, calculated
Squared Age (age_squared)	[441 ; 4,489]	Squared age of respondent, calculated
General Health (health)	[1 ; 5] 1 = Very good ; 2 = Good ; 3 = Fair ; 4 = Bad ; 5 = Very bad	How is your health in general? Would you say it is . . .
Education (educ)	[1 ; 8] 1 = Less than lower secondary ; 2 = Lower secondary ; 3 = Lower tier upper secondary ; 4 = Upper tier upper secondary ; 5 = Advanced vocational ; 6 = Lower tertiary (BA level) ; 7 = Higher tertiary (> MA level) ; 8 = None of the above or other	Highest level of education, ES - ISCED?
Equivalised Household Income in € per Month (equiv_hhinc)	[25 ; 6,125]	Using this card, if you add up the income from all sources, which letter describes your household's total net income? If you don't know the exact figure, please give an estimate. Use the part of the card that you know best: weekly, monthly or annual income (see Appendix A).
Logarithm of Equivalised Household Income in € per Month (log_equiv_hhinc)	[3.22 ; 8.72]	Logarithm of Equivalised Household Income
Belonging to Ethnic Minority (eth_minority)	[0 ; 1] 0 = No ; 1 = Yes	Do you belong to a minority ethnic group in [country]?
Religion (religious)	[0 ; 1] 0 = No ; 1 = Yes	Do you consider yourself as belonging to any particular religion or denomination?
Marital Status (marstat)	[1 ; 6] 1 = Married ; 2 = In a partnership ; 3 = Separated ; 4 = Divorced ; 5 = Widowed ; 6 = Never married or similar	This question is about your legal marital status not about who you may or may not be living with. Which one of the descriptions on this card describes your legal marital status now?
Gender (gender)	[0 ; 1] 0 = Female ; 1 = Male	Respondent's Gender
Occupation (occup8)	[1 ; 8] 1 = Self-employed professionals and large employers ; 2 = Small business owners ; 3 = Technical (semi-)professionals ; 4 = Production workers ; 5 = (Associate) managers ; 6 = Clerks ; 7 = Socio-cultural (semi-)professionals ; 8 = Service workers	What is/was the name or title of your main job? In your main job, what kind of work do/did you do most of the time?

selected for further analysis, as it more accurately captures the actual working hours of the respondents by taking into account the overtime they have worked. This decision is validated in Appendix B (p. 75). Third, the analysis focuses on adults of working age, restricting the age range to 21 to

67, the mandatory retirement age in Iceland, thereby ensuring that only individuals likely fully integrated into the labour force are included and those nearing retirement or still in education are excluded. Furthermore, the occupation variable is refined using the Oesch-class schema (Oesch,

Table 2: Summary Statistics of the Data Sample

SUMMARY STATISTICS									
Summary Statistics									
Variables	Overall			Women			Men		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Happiness (happy)	3,935	8.22	1.44	2,011	8.27	1.39	1,924	8.16	1.5
Weekly Working Hours incl. Overtime (work_hours_ot)	3,574	40.73	14.17	1,842	36.49	13.31	1,732	45.25	13.65
Squared Weekly Working Hours incl. Overtime (work_hours_ot_squared)	3,574	1,859.99	1,120.03	1 842	1,508.90	941.97	1,732	2,233.38	1,172.60
Weekly Working Hours incl. Overtime in Categories (hours_category)	3,960	2.73	1.26	2,022	2.40	1.22	1,938	3.08	1.20
Age (age)	2,970	44.75	13.15	1,527	44.93	13.05	1,443	44.57	13.26
Squared Age (age_squared)	2,970	2,175.64	1,171.74	1,527	2,188.56	1,163.57	1,443	2,161.98	1,180.56
General Health (health)	3,948	1.93	0.88	2,017	1.96	0.90	1,931	1.89	0.86
Education (edu)	3,876	4.23	1.92	1,983	4.34	2.00	1,893	4.12	1.82
Equivalised Household Income in € per Month (equiv_hhinc)	3,552	1,145.10	1,011.18	1,796	1,093.93	978.23	1,756	1,197.43	1,041.5
Logarithm of Equivalised Household Income in € per Month (log_equiv_hhinc)	3,552	6.56	1.14	1,796	6.50	1.16	1,756	6.62	1.12
Belonging to Ethnic Minority (eth_minority)	3,009	0.03	0.17	1,525	0.03	0.17	1,484	0.03	0.17
Religion (religious)	3,935	0.46	0.50	2,007	0.50	0.50	1,928	0.41	0.49
Marital Status (marstat)	2,322	4.75	1.82	1,190	4.73	1.77	1,132	4.77	1.86
Occupation (occup8)	3,844	5.23	2.20	1,969	5.90	2.04	1,875	4.53	2.14

N = Sample Size
SD = Standard Deviation

2006) to condense 594 categories into an 8-class schema that indicates social classes, aiding in exploring how positions within the employment structure influence happiness. Last but not least, the inclusion of squared terms for age and working hours allow for non-linear effects on happiness, such as (inverse) U-shaped relationships as noted in the literature (Blanchflower & Oswald, 2004; Meier & Stutzer, 2004; Rätzl, 2012). Finally, where deemed appropriate (such as for ethnicity, gender, and religion), variables are re-coded to accurately model a binary variable with outcomes that take on only values of 0 and 1. The final sample for estimation therefore includes 15 variables with observations on a total of 3,975 individuals surveyed captured by the data.

Descriptive statistics from Table 2 offer initial insights into the Iceland sample data, segmented by gender. The data

show an equal distribution of men and women with an overall mean age of around 45 years, reporting good health, an average happiness score of 8.22, and about 41 weekly working hours, including overtime. Notably, women work considerably less than their male counterparts (36.49 hours versus 45.25 hours) and earn less monthly (€1,093 versus €1,197), yet they report higher happiness scores (8.27 versus 8.16). This discrepancy in working hours may stem from the fact that women often take on lower-paying part-time jobs or exit the labour market completely (Chung & van der Horst, 2018; Vlasblom & Schippers, 2006), a trend linked to excessive spousal working hours in dual-earner households where men's careers are prioritised, thus limiting women's ability to maximise their potential in the labour market (Cha & Weeden, 2014).

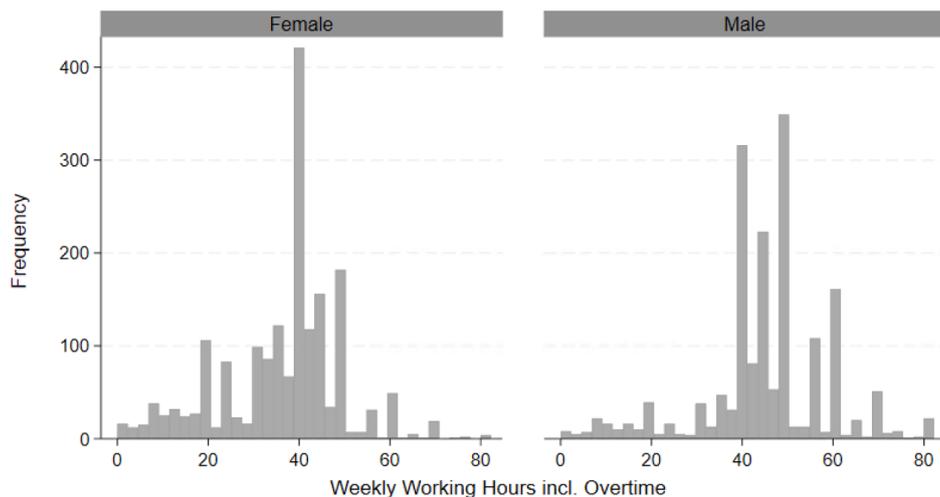


Figure 1: Distribution of Weekly Working Hours by Gender

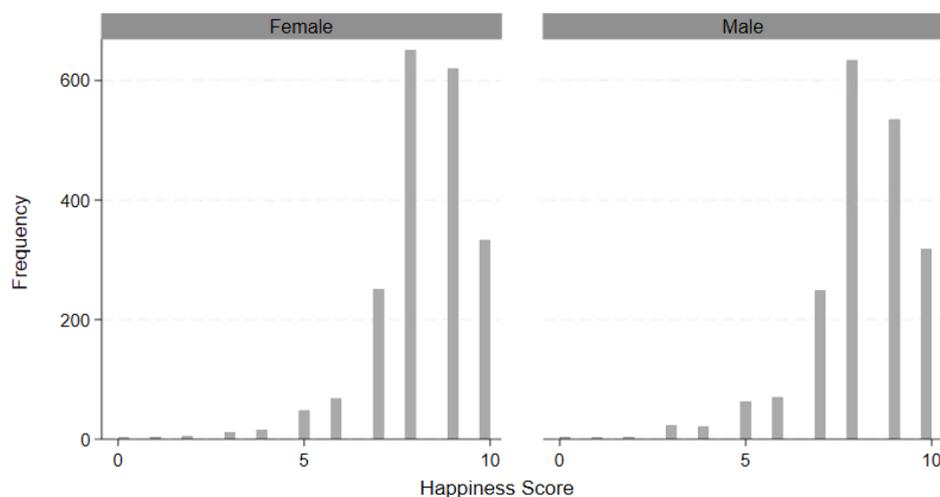


Figure 2: Distribution of Happiness Scores by Gender

Table 3: Key Variables With Over 1% of Values Missing

SUMMARY STATISTICS

Key Variables With Over 1% of Values Missing

Variables	Values Missing (/ 3,975)	Percentage of Values Missing
Marital Status (marstat)	1,642	41.31%
[Squared] Age [squared_age] (age)	988	24.86%
Belonging to Ethnic Minority (eth_minority)	954	24.00%
[Logarithm of] Equivalised Household Income in € per Month [log_equiv_hhinc] (equiv_hhinc)	414	10.42%
[Squared] Weekly Working Hours incl. Overtime [squared_work_hours_ot] (work_hours_ot)	393	9.89%
Occupation (occup8)	120	3.02%
Education (edu)	88	2.21%

Furthermore, Figure 1 illustrates the frequency of working hours by gender, highlighting the remarks made above: The majority of women work in jobs with lower weekly working hours, with most women working a full-time working week of 40 hours, while their male counterparts tend to work longer hours, with frequencies peaking between 40 and 60 hours.

Additionally, Figure 2 shows the distribution of happiness scores in Iceland by gender, demonstrating relative frequencies in the lower extremes (0 = Extremely unhappy) to be very scarce, while happiness scores lean towards the upper end (10 = Extremely happy) for both genders. This reflects Iceland's status as one of the happiest countries worldwide, emphasising its unique position in maintaining high happiness levels even amongst socio-economically disadvantaged groups (Gudmundsdóttir et al., 2016).

Finally, Table 3 displays variables with significant missing values in the dataset, potentially impacting their explanatory power. Due to a high percentage of missing values and its statistical insignificance (see Appendix B, p. 75), the marital status variable is excluded from the final regression analysis. Similarly, the variable indicating belonging to an ethnic minority, which lacks data for the latest survey wave (2020), is also excluded to maintain consistency across all years.

3.2. From Utility Theory to the Analytical Framework for Survey Data Analysis

Utility theory in neoclassical economics utilises the utility concept to interpret individual preferences and decision-making processes, positing that individuals aim to maximise their utility (Fishburn, 1968), defined as satisfaction or well-being. According to this theory, an individual's utility function is denoted as:

$$U = U(Y, L) \quad (1)$$

, depending on income Y , a product of working time H , and the wage rate w , often also called consumption C (Rätzel, 2012), and an individual's leisure time L which is the non-working time within total available time T (Rätzel, 2012), so that:

$$L = T - H \quad (2)$$

Both Y and L enter the utility function positively, subject to diminishing marginal returns with negative second derivatives, hence assuming:

$$U_Y > 0; U_L > 0; U_{YY} < 0; U_{LL} < 0 \quad (3)$$

Literature around SWB, for the most part, models reported SWB r as a function of actual SWB h , with responses typically tied to life satisfaction or happiness (Dolan et al., 2008):

$$SWB = r(h) \quad (4)$$

The utility (1) and SWB (4) function can be combined in a model where SWB is a function of working hours, income, and a vector \mathbf{X} of socio-demographic and socio-economic characteristics affecting happiness, therefore taking on the form of:

$$SWB = f(H, Y, \mathbf{X}) \quad (5)$$

Lastly, in order to conduct an empirical analysis, (5) is modelled as an additive function that can be estimated via OLS, namely:

$$h = \alpha + \beta_1 * H + \beta_2 * H^2 + \beta_3 * Y + \beta_4 * \mathbf{X} + \varepsilon \quad (6)$$

that has happiness h as a dependent variable, reported in ordered categories, further includes a combination of linear and quadratic terms in hours worked H to allow for potential non-linear effects on happiness, income Y , and a vector X of demographic and household characteristics, namely gender, age (squared), health, education, religion, and occupation. Finally, the error term ε accounts for other unobserved individual differences.

After initial descriptive and correlational analysis, an OLS regression explores the association of working hours with happiness, chosen as literature allows for happiness responses to be treated as cardinal measures where the differences in response levels are consistent across the scale (Dolan et al., 2008). Additionally, findings show that OLS outcomes are comparable to those from multinomial or ordered logit models (Blanchflower & Oswald, 2011; Ferrer-i-Carbonell & Frijters, 2004), simplifying the interpretation of regression coefficients to detect the effect of a one unit change in working hours on the scale of reported happiness levels. In order to dive deeper into the working hours - happiness nexus, two separate measures of working hours are employed, first, a continuous measure, and second, working hours are split into six brackets. Thereupon, the analysis first assesses the raw relationship between working hours and happiness, then uses sequential estimation to assess coefficient robustness against control variables, particularly gender and occupational differences. Additionally, post-estimation marginal effects are evaluated to predict the benefits of reduced working, and for the purpose of testing adequacy of variables for modelling happiness and their joint significance, a Wald Test is carried out at a significance level of 5%, facilitating an exploration of the extent to which the data substantiate the proposed hypotheses.

A robustness check is performed by using the same variables and procedures on the ESS dataset from France, chosen because of its comparable significant reform to Iceland's FDWW experiment. In the year 2000, under Jacques Chirac's presidency (1995-2007), prime minister Lionel Jospin reduced the workweek from 40 to 35 hours without reducing

workers' pay, aiming to recover from a national recession and boost job creation (Ashta, 2000). Ultimately, this reform enhanced employee's well-being and work-life balance (Hayden, 2006), making France a suitable comparison for validating the empirical results.

Finally, the unique nature of the ESS dataset necessitates using the correct estimation procedure that considers weighting, clustering, and stratification when analysing survey data. The *svyset* function in Stata is employed to declare the survey design, incorporating primary sampling units (psu) for cluster sampling, stratification (*stratum*) to handle representativeness of different population groups, and weights (constructed from the design weight with post-stratification and population size adjustments) to balance selection probabilities and address non-response (Kaminska & Lynn, 2017). This methodological consideration is crucial for upholding the validity and reliability of the conclusions drawn from the ESS data, as it helps mitigate issues like coverage, sample attrition, and non-response errors, ensuring that the analysis produces representative and unbiased estimates (ESS, 2024d). Chapter 4 explores the results of this analysis.

4. Results

In order to get a first idea of how the variables and especially weekly working hours are related to happiness, a weighted Pearson's pairwise correlations test is run. There is a very small positive correlation between happiness and weekly working hours (squared), $r(3,565) = 0.021$ (0.016), $p > 0.05$, with working hours explaining 0.04 (0.03)% of the variation in happiness scores. Weighted pairwise correlations are also reported for all other variables (see Appendix C, p. 79). The table reveals (i) most variables as significantly correlated at the 5% level, and (ii), all variables including working hours and squared working hours to be positively correlated with happiness, except for gender, health, and occupation that exhibit a negative correlation with happiness.

Below descriptive statistics and estimated regression models will give more insights into the suggested positively related relationship of working hours and happiness, any potential non-linearities, and their association within the FDWW experiment.

4.1. Non-Linear Relationship Test

To begin with, simple descriptive statistics across genders in Table 4 portray neither a negative nor an inverse U-shaped relationship between working hours and happiness. Instead, weekly working hours seem to be positively related to happiness up until 60 hours of weekly working time, after which each additional hour of work is linked to decreased happiness. Most employed Icelanders work 41-50 hours weekly, with women typically working less than 40 hours, while men are featured in the category of excessive working hours (60+). In particular, women's average happiness is highest in jobs working less than 40 hours, whereas

men only reach their highest average happiness in jobs working more than 40 hours. Further, women's average happiness stays relatively constant when working between 1 and 60 hours a week, and decreases when working 60+ hours. For men, an important exception to their average happiness is observed in jobs with very low weekly working hours (1-30 hours), illustrating their aversion towards part-time jobs or jobs with low working hours. Lastly, the results on happiness scores allow consideration of the non-working population which is recorded at zero weekly working hours. Indeed, women report a much higher happiness score at zero working hours than men, findings that can be mainly attributed to traditional gender roles, in which women staying at home are considered as housewives, while for men not being at work represents unemployment (Rudolf, 2013).

Table 5 presents the regression results from OLS testing where model 1 treats weekly working hours as a linear, categorical variable and model 2 introduces a squared term to explore non-linear effects. It appears that longer weekly working hours are generally associated with a higher level of happiness, precisely, a 0.012 point increase on the ten-point happiness scale per additional hour worked, although the quadratic term's negligible diminishing rate and insignificant coefficients suggest minor impact. Furthermore, model 7 highlights income as a significant factor, with the impact of working hours dropping to a third in size when controlled for. At first glance these findings therefore appear to contrast previous findings of a negative or inverse U-shaped relationship, making it of interest to delve further into the worktime - happiness relationship. Model 3 does this by breaking out weekly hours into six brackets, relative to being unemployed (0 hours), revealing all effects to be statistically significant at $p < 0.05$ and all working-hour brackets to exhibit a negative relationship with happiness. Effects are the most negative at the margins (1-30 hours and 60+ hours) and become the least negative in the 41-50 hour bracket, thereby indicating an inverse U-shaped relationship where working longer or shorter than a standard workweek is associated with increased diminishing worker happiness. Figure 3 visually illustrates this relationship in a boxplot. Adding controls (models 4 to 6), however, seems to undo this effect, as all worktime coefficients lose their statistical significance at the conventional levels, change signs to being positive, and are notably smaller in size. In particular, increased happiness scores are observed with increasing weekly working hours, that is, the highest happiness is achieved when working an additional hour in the 50-60 and 60+ hours bracket, holding all other factors constant (respectively, an increase of 0.191 and 0.140 points on the ten-point happiness scale). Nonetheless, despite the generally positive association of working hours, the effects are very small and therefore results can only be evaluated very close to zero, indicating that the effects are not to a degree that is large enough for a significant change in working hours to have a detectable effect on happiness scores once people are employed.

Table 5 — continued

<i>Small Business Owners</i>						-0.267*	-0.396***	-0.257*	-0.363**
						(0.137)	(0.137)	(0.141)	(0.142)
<i>Technical (Semi-)Professionals</i>						-0.129	-0.189	-0.116	-0.147
						(0.134)	(0.130)	(0.135)	(0.131)
<i>Production Workers</i>						-0.438***	-0.411***	-0.413***	-0.339**
						(0.142)	(0.150)	(0.144)	(0.151)
<i>(Associate) Managers</i>						-0.099	-0.237**	-0.073	-0.221*
						(0.122)	(0.119)	(0.121)	(0.120)
<i>Clerks</i>						-0.377**	-0.355**	-0.388**	-0.329**
						(0.167)	(0.157)	(0.172)	(0.161)
<i>Socio-Cultural (Semi-)Professionals</i>						-0.134	-0.243**	-0.113	-0.217*
						(0.121)	(0.119)	(0.121)	(0.122)
<i>Service Workers</i>						-0.288**	-0.269**	-0.294**	-0.205
						(0.127)	(0.129)	(0.127)	(0.131)
Gender (gender) <i>Female</i>							-0.209***		-0.232***
<i>Male</i>							(0.066)		(0.069)
Age (age)							0.005		0.007
							(0.157)		(0.016)
Squared Age (age_squared)							0.000		0.000
							(0.000)		(0.000)
General Health (health) <i>Very Good</i>									
<i>Good</i>							-0.386***		-0.380***
							(0.050)		(0.508)
<i>Fair</i>							-1.085***		-1.108***
							(0.100)		(0.108)
<i>Bad</i>							-1.635***		-1.675***
							(0.250)		(0.262)
<i>Very Bad</i>							-2.77***		-3.149***
							(0.709)		(0.725)
Education (edu) <i>Less Than Lower Secondary</i>									
<i>Lower Secondary</i>							-0.009		-0.005
							(0.473)		(0.485)
<i>Lower Tier Upper Secondary</i>							0.130		0.040
							(0.475)		(0.487)
<i>Upper Tier Upper Secondary</i>							-0.064		-0.069
							(0.476)		(0.489)
<i>Advanced Vocational</i>							-0.114		-0.152
							(0.474)		(0.484)
<i>Lower Tertiary (BA Level)</i>							-0.069		-0.075
							(0.472)		(0.483)
<i>Higher Tertiary (> MA Level)</i>							-0.269		-0.245
							(0.473)		(0.485)
<i>None of the Above or Other</i>							0.890		0.675
							(0.585)		(0.671)
Religion (religious) <i>No</i>							0.204***		0.171***
<i>Yes</i>							(0.058)		(0.060)
_cons	7.999***	7.871***	8.951***	7.410***	7.782***	7.321***		7.058***	7.331***
	(0.096)	(0.149)	(0.315)	(0.415)	(0.460)	(0.643)		(0.268)	(0.631)
Observations	3,085	3,085	3,383	3,065	3,023	2,305	2,838	2,814	2,179
R-Squared	0.001	0.002	0.004	0.033	0.040	0.178	0.031	0.038	0.180
Prob > F	0.135	0.161	0.007	0.000	0.000	0.000	0.000	0.000	0.000

Robust Standard Errors in Parantheses
P > |t| = Result of the Two-Tailed t-Test
Significance Levels: * p < 0.1 ; ** p < 0.05 ; *** p < 0.01

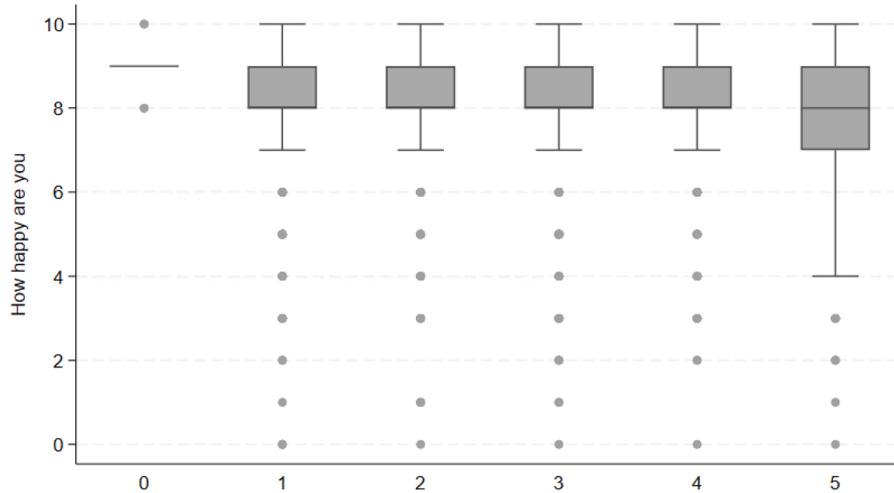


Figure 3: Relationship Between Weekly Working Hours Categories and Happiness

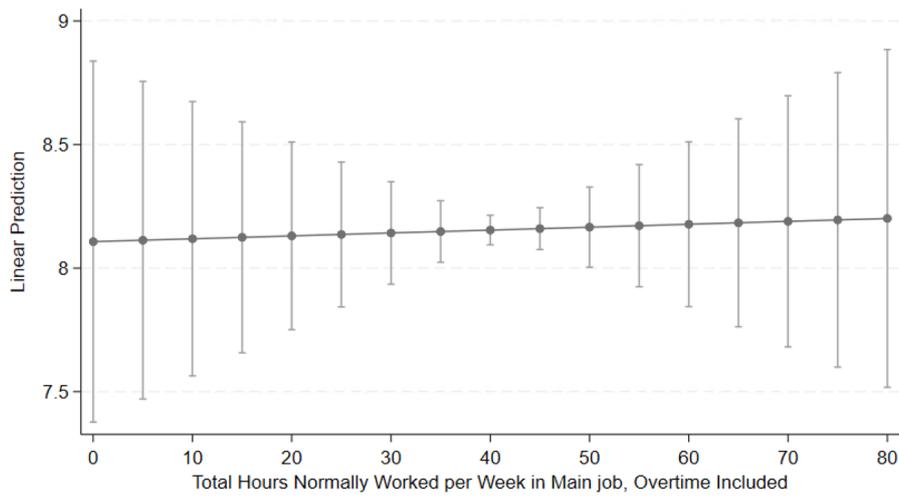


Figure 4: Predictive Margins of Continuous Working Hours With 95% Confidence Intervals

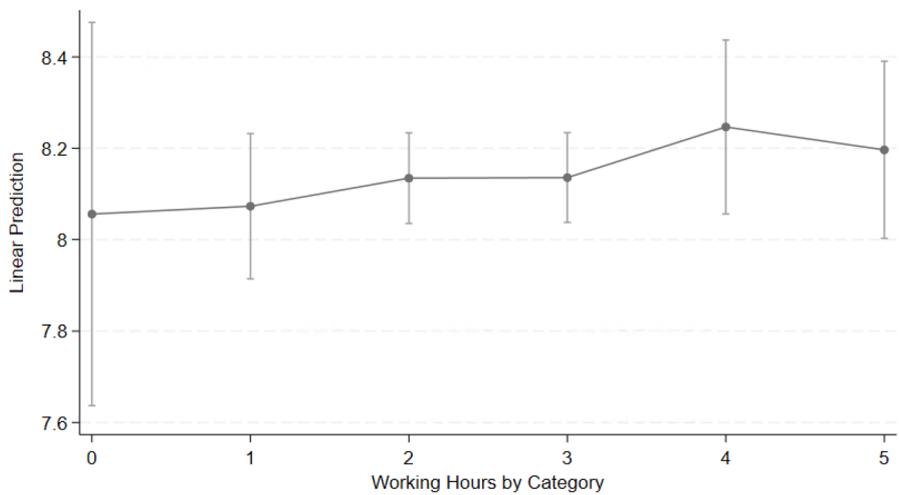


Figure 5: Predictive Margins of Working Hours by Category With 95% Confidence Intervals

Last but not least, being a man and in poorer health are linked to reduced happiness scores, whereas religiousness enhances it, all statistically significant at $p < 0.01$. Finally, employment in upper-middle-class jobs, the baseline of the occupation variable, correlates with the highest happiness, contrasting with lower happiness in the working class (production workers) and lower-middle and salaried class roles. Especially, accounting for occupations reveals a diminished link between happiness, implying that income's positive effect on happiness is more pronounced across different occupations than within them.

Figure 3 illustrates the relationship between categories of weekly working hours and happiness as described by model 3 above, suggesting the extremes of the working hours spectrum (very low and very high hours) to be significantly associated with lower happiness, indicating an inverse U-shaped relationship, while the mean happiness over other categories (especially in category 2 and 3) remains stable.

Furthermore, margins estimation shows minimal variation in predicted happiness scores with increases in working hours once other factors are controlled for, especially between the 30- and 40-hour marks (Figure 4) and across the second and third working hour categories (Figure 5), ranges typically corresponding to the scope of a FDWW experiment. The flat trend suggests that weekly working hours alone do not significantly impact predicted happiness, indicating that other factors may play a more decisive role in shaping the overall happiness of the Icelandic workforce.

The literature establishes happiness factors to differ by gender (Alesina et al., 2004; Louis & Zhao, 2002), prompting separate model estimations for men and women, detailed in Appendix D, page 80. Results indicate a positive effect of weekly working hours on men's happiness and a negative effect for women, though neither is statistically significant. Specifically, men's slight positive coefficient in happiness is driven entirely by the 1-30 hours working bracket, where each additional hour lowers happiness by 0.145 points on the ten-point happiness scale, highlighting men's preference for full-time work over part-time or unemployment, possibly due to the negative stigma of underemployment, and suggesting that men's part-time work is due to involuntary underemployment rather than a choice (Rudolf, 2013). Conversely, income appears to play a greater role for women. Nevertheless, similar to the complete model, these estimations are also very close to zero, hence this regression output also reveals that the estimated effects of weekly working hours for both genders are not large enough so that once employed a significant change in working hours can be detected.

Finally, as the FDWW experiment has mainly targeted public sectors, breaking out the hours - happiness relationship by the eight major occupational classifications allows to shed more light on nuances in this relationship. Appendix E, p. 82 reveals that with the exception of small business owners and technical (semi-) professionals, working and lower-middle class workers generally show a negative relationship between working hours and happiness, whereas workers in the salaried middle and traditional upper-middle classes

display a positive relationship, suggesting that higher occupational classes find more fulfilment in increased working hours. Additionally, income matters differently and more for certain occupations, being more pronounced in less skilled or educational roles (as also highlighted by Golden and Okulicz-Kozaryn, 2015), evident from the higher income coefficient for clerks compared to professionals or managers. Yet, like previous findings, these evaluations are very close to zero and working-hour results are statistically insignificant at the conventional levels, implying no detectable effect in the worktime - happiness nexus when differentiating between occupations.

4.2. Results in the Context of the Four-Day Working Week

After focusing on general results, this section turns towards the FDWW in the Icelandic context, conducting an evaluation of evidence based for policy with the state and quality of data that is available.

By taking a closer look at the average happiness reported before, during, and after the FDWW experiment, it can be detailed how the distribution of working hours and happiness scores has shifted for women and men. Table 6 shows that during the experiment, overall, there are 5% more people working lower hours (31-40 hours), however, with a lower happiness score as well. For women, especially, a shift from not working at all (0 hours), working part-time (1-30 hours), and working 60+ hours towards working full-time hours (31-50 hours) is observable, with the happiness score fluctuating depending on the working hours category, but decreasing overall. For men, the same pattern applies, that is, the share of working 31-50 hours per week increases by around 4% for both gender groups, while the happiness score decreases. After the FDWW experiment, this trend continues, with, overall, more people working full-time (31-50 hours) and fewer working excessive hours (60+ hours) than before the experiment. Nonetheless, a consistently decreasing happiness score is reported over the years, making it of importance to analyse whether the low happiness scores can really be attributed to (reduced) working hours, or rather to other factors.

An OLS regression is run on the data from before, during, and after the FDWW trial to thoroughly evaluate the experiment, however, as the data is not correlated with the treatment, a true before-after evaluation cannot be made, necessitating cautious assessment of results (more about this limitation is discussed in chapter 5.1). Regression results in Table 7 provide information about the worktime - happiness relationship, indicating the worktime coefficients to be statistically insignificant at all conventional levels, as well as positive before and after, and negative during the trial, implying that with the trial, any additional hour of work was perceived as more stressful due to disruptions in usual work routines, increased workload, or poor implementation of the policy. However, this effect is so small that it should be neglected, the results rather underline again that the coefficients of working hours have no detectable effect on happiness in the periods

Table 6: Average Happiness by Weekly Working Hours and Gender - FDWW Experiment

DESCRIPTIVE STATISTICS

Average Happiness by Weekly Working Hours and Gender Before, During, and After the FDWW Experiment

	Overall					
	Before		During		After	
	N [%]	Average Happiness	N [%]	Average Happiness	N [%]	Average Happiness
Hours 0	5 [0.38%]	9.20	4 [0.23%]	8.75	–	–
Hours 1-30	259 [19.59%]	8.34	310 [17.87%]	8.10	148 [16.40%]	8.05
Hours 31-40	295 [22.32%]	8.28	480 [27.67%]	8.19	301 [33.33%]	8.13
Hours 41-50	397 [30.03%]	8.39	571 [32.91%]	8.25	258 [28.57%]	8.16
Hours 51-60	139 [10.51%]	8.37	184 [10.60%]	8.35	73 [8.08%]	8.01
Hours 60+	227 [17.17%]	8.22	186 [10.72%]	8.16	123 [13.62%]	7.92
Total	1,322 [100%]	8.33	1,735 [100%]	8.21	903 [100%]	8.09

	Women					
	Before		During		After	
	N [%]	Average Happiness	N [%]	Average Happiness	N [%]	Average Happiness
Hours 0	4 [0.30%]	9.25	3 [0.17%]	9.00	–	–
Hours 1-30	196 [14.83%]	8.48	217 [12.51%]	8.18	105 [11.64%]	8.15
Hours 31-40	202 [15.28%]	8.33	295 [17.00%]	8.34	188 [20.82%]	8.18
Hours 41-50	154 [11.65%]	8.41	240 [13.83%]	8.29	110 [12.18%]	7.92
Hours 51-60	29 [2.19%]	8.10	44 [2.53%]	8.34	22 [2.43%]	8.41
Hours 60+	89 [6.73%]	8.40	81 [4.67%]	8.25	43 [4.76%]	7.71
Total	674 [50.98%]	8.40	880 [50.72%]	8.28	468 [51.83%]	8.08

N = Sample Size

Happiness Scale: 0-10 with 0 = "Extremely Unhappy" and 10 = "Extremely Happy"

Before = Years 2004, 2012 ; During = Years 2016, 2018 ; After = Year 2020

% from overall total in brackets

evaluated and the FDWW thus does not seem to have significantly affected overall happiness of employees in Iceland.

A proposed FDWW experiment should aim to provide at least some correlational evidence to validate its effectiveness

and implementation, and while there is evidence from the descriptive statistics that those working extreme hours (60+) may experience higher happiness scores when transitioning to fewer working hours, there is no evidence that the FDWW

Table 6 — continued

	Men					
	Before		During		After	
	N [%]	Average Happiness	N [%]	Average Happiness	N [%]	Average Happiness
Hours 0	1 [0.08%]	9	1 [0.06%]	8	–	–
Hours 1-30	63 [4.76%]	7.90	93 [5.36%]	7.91	43 [4.76%]	7.79
Hours 31-40	93 [7.04%]	8.17	185 [10.67%]	7.93	113 [12.51%]	8.04
Hours 41-50	243 [18.38%]	8.38	331 [19.08%]	8.22	148 [16.39%]	8.34
Hours 51-60	110 [8.32%]	8.44	140 [8.07%]	8.35	51 [5.65%]	7.84
Hours 60+	138 [10.44%]	8.11	105 [6.05%]	8.09	80 [8.86%]	8.03
Total	648 [49.02%]	8.26	855 [49.28%]	8.13	435 [48.17%]	8.09

Table 7: Regression Output - Effect of Weekly Working Hours on Happiness Before, During, and After the FDWW Experiment

REGRESSION OUTPUT

Survey Data Regression Output Before, During, and After the FDWW Experiment

Variables	Strata with single sampling unit treated as certainty units					
	Before		During		After	
	(1)	(2)	(3)	(4)	(5)	(6)
	Happiness (happy)					
Weekly Working Hours incl. Overtime (work_hours_ot)	0.167 (0.017)		-0.011 (0.011)		0.013 (0.021)	
Squared Weekly Working Hours incl. Overtime (work_hours_ot_squared)	-0.000 (0.000)		0.000 (0.000)		-0.000 (0.000)	
Weekly Working Hours incl. Overtime in Categories (hours_category) <i>Hours 0</i>						
<i>Hours 1-30</i>		-0.358 (0.340)		0.192 (0.161)		
<i>Hours 31-40</i>		-0.122 (0.303)		0.249 (0.148)		-0.037 (0.192)
<i>Hours 41-50</i>		-0.239 (0.291)		0.229 (0.153)		0.135 (0.198)
<i>Hours 51-60</i>		-0.308 (0.342)		0.511 (0.204)		0.055 (0.293)
<i>Hours 60+</i>		-0.244 (0.359)		0.246 (0.188)		0.234 (0.281)
Logarithm of Equivalised Household Income in € per Month (log_equiv_hhinc)	0.102* (0.586)	0.134** (0.053)	0.238*** (0.050)	0.251*** (0.048)	0.246*** (0.068)	0.236*** (0.067)
Occupation (occup8) <i>Self-Employed Professionals and Large Employers</i>						
<i>Small Business Owners</i>	-0.426 (0.327)	-0.370 (0.330)	-0.366* (0.194)	-0.311* (0.181)	-0.212 (0.248)	-0.296 (0.248)
<i>Technical (Semi-)Professionals</i>	-0.382 (0.268)	-0.369 (0.295)	-0.191 (0.180)	-0.140 (0.175)	0.189 (0.175)	0.149 (0.203)

Table 7 — continued

<i>Production Workers</i>	-0.002 (0.363)	-0.035 (0.358)	-0.403** (0.200)	-0.439** (0.195)	-0.477** (0.206)	-0.605** (0.240)
<i>(Associate) Managers</i>	-0.342 (0.283)	-0.270 (0.295)	-0.297* (0.159)	-0.260* (0.154)	0.071 (0.179)	0.067 (0.191)
<i>Clerks</i>	-0.412 (0.337)	-0.281 (0.335)	-0.548** (0.232)	-0.546** (0.230)	0.099 (0.256)	0.103 (0.258)
<i>Socio-Cultural (Semi-)Professionals</i>	-0.486* (0.283)	-0.420 (0.285)	-0.201 (0.172)	-0.172 (0.163)	-0.004 (0.165)	-0.026 (0.185)
<i>Service Workers</i>	-0.554* (0.293)	-0.454 (0.300)	-0.160 (0.182)	-0.213 (0.176)	-0.023 (0.167)	-0.030 (0.195)
Gender (gender) <i>Female</i>	-0.498** (0.172)	-0.445** (0.174)	-0.252** (0.090)	-0.250** (0.086)	-0.039 (0.103)	-0.009 (0.107)
<i>Male</i>						
Age (age)	-0.036 (0.039)	-0.029 (0.038)	0.023 (0.021)	0.014 (0.020)	0.023 (0.025)	0.022 (0.025)
Squared Age (age_squared)	0.001 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
General Health (health) <i>Very Good</i>						
<i>Good</i>	-0.391** (0.126)	-0.376** (0.123)	-0.330** (0.068)	-0.352** (0.066)	-0.484** (0.093)	-0.471** (0.097)
<i>Fair</i>	-1.152** (0.211)	-1.204** (0.187)	-0.969** (0.158)	-0.936** (0.143)	-1.384** (0.201)	-1.339** (0.199)
<i>Bad</i>	-2.001** (0.609)	-1.94** (0.625)	-1.495** (0.366)	-1.507** (0.317)	-1.627** (0.397)	-1.592** (0.457)
<i>Very Bad</i>	-3.587** (1.421)	-3.530** (1.368)	-3.279** (1.126)	-2.385** (1.000)	-2.518** (0.921)	-2.464** (0.918)
Education (edu) <i>Less Than Lower Secondary</i>						
<i>Lower Secondary</i>	-0.314 (0.731)	-0.411 (0.738)	0.186 (0.574)	0.155 (0.577)	0.320 (0.458)	0.269 (0.502)
<i>Lower Tier Upper Secondary</i>	-0.115 (0.717)	-0.251 (0.721)	0.167 (0.585)	0.051 (0.587)	0.346 (0.469)	0.391 (0.517)
<i>Upper Tier Upper Secondary</i>	0.067 (0.762)	-0.102 (0.765)	0.126 (0.583)	0.100 (0.585)	0.041 (0.467)	-0.003 (0.514)
<i>Advanced Vocational</i>	-0.272 (0.716)	-0.335 (0.723)	0.056 (0.577)	0.049 (0.579)	0.039 (0.478)	-0.037 (0.516)
<i>Lower Tertiary (BA Level)</i>	-0.173 (0.738)	-0.250 (0.740)	0.100 (0.577)	0.018 (0.581)	0.130 (0.441)	0.172 (0.488)
<i>Higher Tertiary (> MA Level)</i>	-0.023 (0.724)	-0.321 (0.735)	-0.076 (0.581)	-0.160 (0.587)	-0.118 (0.459)	-0.166 (0.505)
<i>None of the Above or Other</i>			0.810 (0.733)	0.973 (0.673)		
Religion (religious) <i>No</i>	0.169 (0.143)	0.154 (0.140)	0.154* (0.082)	0.206** (0.075)	0.120 (0.095)	0.147 (0.096)
<i>Yes</i>						
_cons	8.744** (1.214)	9.164** (1.283)	6.649** (0.778)	6.395** (0.760)	5.775** (0.779)	6.189** (0.828)
Observations	468	502	1,155	1,218	556	585
R-Squared	0.212	0.209	0.176	0.179	0.231	0.228
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000

Robust Standard Errors in Parantheses

P > |t| = Result of the Two-Tailed t-Test

Significance Levels: * p < 0.1 ; ** p < 0.05 ; *** p < 0.01

Before = Years 2004, 2012 ; During = Years 2016, 2018 ; After = Year 2020

universally improves happiness scores across different working hours. To explore this finding further, a two-sample t-test is conducted to compare differences in means between those

working 35 hours or less and those working more than 35 hours. These two groups are selected based on the concept of the FDWW, which aims to reduce working hours to less

than a full-time 36-40 hour week, thus allowing the impact of reduced working hours as proposed by the FDWW to be examined. The output of the t-test compares the mean happiness scores between these two groups and reveals that the mean happiness of those working less than 35 hours per week ($M = 8.250$, $SD = 1.471$) compared to those working more than 35 hours ($M = 8.227$, $SD = 1.421$) is not significantly different, $t(3117) = 0.141$, $p = 0.888$ (see Appendix F, p. 85). These results indicate mean happiness scores to be stable and small changes observed in happiness scores during the trial to not be substantial enough to be considered significant, supporting the conclusion that the FDWW does not significantly impact overall happiness and individuals working a potential FDWW are just as happy as those working longer hours.

4.3. Robustness Check

Robustness checks verify the sensitivity of above results to specifications made, ensuring validity and applicability of findings. Firstly, Wald Tests on Icelandic data indicate no significant relationship between working hours and happiness, confirming that the working hours coefficient may not be interpreted as the true causal effect of happiness (see Appendix G, p. 86). Thereby, both hypothesis H_1 and H_2 cannot be confirmed, as neither an inverse U-shaped relationship, nor a shifting association throughout the phases of the FDWW trial is detected. Similarly, no further specifications (gender or occupation) yield a consistent link between working hours and happiness. Furthermore, regressions are estimated using the heteroskedastic-robust version of standard errors to provide unbiased standard error estimates, accounting for the variances in error terms. Finally, a robustness check by testing the model under varying socio-economic and cultural conditions using French data corroborates the methodological soundness of the analysis (see Appendix G, p. 86). Despite cultural and policy differences influencing outcomes, such as the insignificance of religion in France, the foundational models remain valid, highlighting how in another affluent country the influence of working hours on happiness remains similarly non-significant, reinforcing the broader applicability of these findings across economically developed contexts.

5. Discussion

Now that all results are presented, this section delves into a discussion, which, to provide clarity, is organised into three main sections: an evaluation of the methodological approach and data used in the analysis, a discussion of the results, contextualised within the existing literature, and an identification of suggested avenues for future research, concluding with perspectives on the broader implications of this research topic.

5.1. Discussion of the Approach

The methodological approach of this thesis, which employs OLS regressions to analyse ESS data, is pivotal to

methodically address the impact of the FDWW in Iceland, leveraging a robust dataset to scrutinise the interplay between working hours and employee happiness. Despite its strengths, this approach is not without limitations that merit consideration.

A critical drawback of above analysis lies in the lack of correlation between the dataset used and the actual treatment applied during the FDWW trial, allowing for no causal interpretation of results, but only an assessment of the direction, form, and strength of the relationship. This disconnection stems primarily from the observational nature of ESS data, which, unlike controlled experimental data, does not establish a causative link between the reduced working hours implemented during the FDWW and subsequent changes in happiness levels. As the ESS dataset does not originate from the controlled experimental setup where participants are randomly assigned to treatment and control groups, it is challenging to attribute observed changes in happiness directly to the FDWW policy implementation, meaning that before-and-after comparisons do not only capture the effects of the policy, but also the influence of other uncontrolled external factors occurring simultaneously.

Furthermore, a significant challenge caused by reverse causality is the potential endogeneity of working hours, a common issue in studies of this nature. Individuals may adjust their working hours in response to their levels of satisfaction, that is, people may allocate more time towards activities they find fulfilling, which means that changes in working hours could be a response to prior dissatisfaction (Rothbard & Edwards, 2003). Consequently, this introduces the challenge of causal relationships explored being biased as workers adjust their hours to enhance personal happiness or alleviate discontent. As the data in this analysis, however, does not allow for any causal inference, this limitation is acknowledged, but does not necessitate changes in the methodological approach undertaken and, additionally, controls are introduced for various factors that might influence both happiness and working hours, such as income and occupation, helping to isolate the effect of working hours from other variables that may confound the results.

Last but not least, the scope of the dataset, primarily encompassing employees from service sectors such as the Reykjavik City Council and Icelandic Government, introduces another limitation. The exclusion of significant economic sectors like manufacturing, construction, or fisheries narrows the generalisability of the findings across Iceland's diverse labour market. This selection bias could skew the results, as the omitted sectors might exhibit different dynamics in the worktime - happiness relationship.

In sum, while the methodological framework of this thesis has limitations inherent to the study of working hours and happiness, their transparent acknowledgement and reporting provides a solid foundation and valuable insights into the association of working hours with happiness, highlighting the need for further research to fully understand and validate the FDWW's impact on workplace happiness.

5.2. Discussion of Results

Central to the discussion of the analysis of working hours and happiness within the Icelandic FDWW trial is whether working hours are a significant determinant of happiness. Results unveil that there is only a minimal effect of working hours on happiness, as once employment is secured, the number of hours worked per week has no effect on happiness, at least not to a degree that it can be detected, thereby rejecting Hypothesis 1. Additionally, the evidence-based assessment of the FDWW trial reveals no significant effect on happiness levels before, during, and after the FDWW, showing that there was a lack of empirical evidence supporting the implementation of the FDWW to begin with, and results indicate that the situation remains unchanged after, that is, the evidence necessary to substantiate the policy's effectiveness in enhancing worker happiness is still lacking. Hypothesis 2 is therefore not supported by the analysis, an outcome not intended by policy makers when introducing a FDWW trial, however, aligning with existing literature that critiques the efficacy of FDWW trials and raises concerns about their foundational assumptions (T. T. Campbell, 2023).

These main results therefore suggest two primary interpretations regarding the non-detectable relationship between working hours and happiness, namely that (i) working hours may not be a substantial determinant of happiness in the context of Iceland, or (ii) there are possibly unobserved factors connected to changes in working hours that may influence happiness scores and are not represented in the model.

The first interpretation seems reasonable in this context, as Iceland provides a unique socio-economic and cultural setting. This includes a strong social welfare system and benefits, such as substantial parental leave, childcare support, and a focus on work-life balance, as well as advanced healthcare policies, low unemployment rates, high standards of living, and a strong community sense that promotes personal fulfilment (Eydal & Gíslason, 2014), which may mitigate the psychological effects of work-related stress differently and better than in other countries and environments. The results from Iceland suggest that merely reducing working hours does not have a substantial direct impact on happiness, which has important implications for labour policies, calling into question the cost-effectiveness of a widespread implementation of FDWW policies. Indeed, while affluent nations may be able to afford the luxury of shorter workweeks without sacrificing economic output, the policy requires a significant allocation of financial resources that could potentially be directed towards other critical sectors. Additionally, with the FDWW trial being limited in scope, covering only 1.3% of Iceland's population, suffering from methodological shortcomings such as lack of data variation that does not allow for treatment effects to be detected, and focusing on reducing weekly hours from 40 to 36 or 35 rather than specifically implementing a four-day week (Kobie, 2021), the interpretation of results is complicated with no evidence available after the analysis that supports the FDWW's efficacy. Consequently, this suggests that the excitement around the FDWW

may be more a product of media hype and the agenda of specific advocacy groups rather than a reflection of a legitimate, evidence-based policy shift that universally benefits the workforce. This reasoning is underscored in particular when taking into consideration that the Icelandic trial has been propelled by an Icelandic non-profit organisation, the Association for Sustainability and Democracy, and the UK think tank Autonomy, both of which were not directly involved in the actual trials, raising concerns about potential biases influencing the public and policy narrative (Kobie, 2021), which emphasises the need for a critical assessment of the purported benefits and the legitimacy of the widespread enthusiasm surrounding the FDWW. Furthermore, regarding the theoretical frameworks introduced in chapter 2, these findings have important theoretical implications for standard economic labour supply models. In particular, the disutility of labour theory, which suggests that leisure inherently provides greater utility than labour, is called into question, as the results do not support a straightforward increase in happiness with reduced working hours, thus challenging the traditional view that leisure alone boosts happiness. Therefore, these results highlight the possible need to re-evaluate traditional models that allow for worker well-being to be included in labour supply decisions (Bakker & Demerouti, 2007). While work may carry a disutility at the margins, it can also contribute positively to happiness (Rätzel, 2012), suggesting a dualistic influence that aligns with and integrates both standard neoclassical theory of individual labour supply and emerging research in the field of happiness economics.

The second interpretation suggests that there are unobserved variables not measured by the model that are connected to changes in working hours and could be influencing happiness levels. In particular, job satisfaction and workplace autonomy play a crucial role in Iceland, where job security and favourable working conditions are prevalent (Gallicie, 2013), potentially making happiness less dependent on working hours. Additionally, social inclusion, a social safety net, trust, peace, and equality are confounding factors that shift along with working hours and influence happiness in Iceland beyond the quantitative measures of work time (Gudmundsdóttir, 2013). As these factors are not directly measured in this thesis, working hours may therefore not appear to be the primary factor influencing happiness based on the available data. The analysis, however, does find gender, health, religion, and income to be significant determinants of happiness and thereby echoes findings in literature. Namely, (i) women exhibit higher happiness, often attributed to societal and psychological factors such as expectations and social roles that influence self-reported happiness measures (Alesina et al., 2004), (ii) good health is strongly linked to higher happiness levels, possibly due to the direct impact of health on daily functioning and overall happiness (Dolan et al., 2008; Liu et al., 2015), (iii) religion provides strong social support networks and coping mechanisms for dealing with life's challenges, enhancing happiness (Clark & Lelkes, 2006; Frey, 2018), and (iv) income significantly reduces the effect

of working hours on happiness, indicating that the happiness associated with more extended work hours is heavily dependent on the income generated rather than the hours themselves. Here, controlling for income modifies the impact of working hours on happiness, suggesting that financial compensation associated with longer working hours might play a more significant role in influencing happiness than previously acknowledged. Such insights are essential in contexts like Iceland, where the baseline levels of happiness and income are already high (Uchida & Rappleye, 2023), implying that the incremental happiness derived from additional income might be minimal, reflecting diminishing returns on happiness as denoted by Easterlin (Easterlin, 1974; Easterlin & O'Connor, 2020). Additionally, while longer working hours are not found to be direct drivers of increased happiness, happiness may be increased by gains in status attributed to either higher relative income or enhanced prestige within the workplace (Altman, 2005; Frey & Stutzer, 2002), with the accompanying income having the potential to offset potential disutilities of labour.

Furthermore, the broader implications of these findings gain significance, especially in light of Greece's recent move to a six-day work week as of July 1st, 2024 (oikonomia, 2024), which contrasts sharply with the trend towards reducing working hours that has gained prominence until now. Greece's recent extension of the workweek aims at addressing economic efficiency rather than worker well-being directly and the intention to increase productivity by adding a sixth workday with compensation incentives underscores the complexity of balancing work hours with productivity and worker satisfaction in different cultural and economic settings. Having observed a diminished positive impact of longer working hours on happiness when controlling for income is a crucial result as Greece implements higher compensation for additional workdays, potentially offsetting the disutility of labour with financial benefits. This serves as a useful counterpoint to Iceland's approach, emphasising the need for policies to consider influencing happiness not directly through reduced hours but through improved economic stability, and inviting policymakers to consider not only the direct effects of reduced working hours on happiness, but also the broader economic and social contexts in which these policies are implemented.

To conclude, the evaluation of the evidence-based policy highlights that (i) working hours have no significant effect on overall happiness, and (ii) the FDWW does not seem to have significantly affected the happiness scores of employees in Iceland. The juxtaposition of the six-day working week in Greece against the Icelandic FDWW experiment further underlines the complexity of applying a one-size-fits-all solution to labour policies and the necessity of tailoring interventions to specific national contexts. While the FDWW offers a promising avenue for wealthy countries to improve work-life balance, the proposed benefits in terms of increased happiness remain without evidence, calling for more rigorous future research to thoroughly examine the impact of working hours on happiness from a controlled experimental setup.

5.3. Suggestions for Further Research, Potential, and Outlook

Further research opportunities in the field of (reduced) working hours and their impact on happiness emerge from synthesising the gaps identified in existing studies, with some already evident from the limitations outlined previously. Given the scarcity of this research topic and the concurrent increase in interest in the FDWW, there are numerous areas for future studies to explore and analyse more comprehensively that may provide new insights and validate the findings of this thesis.

First, future studies should aim to utilise longitudinal data to better capture the long-term effects of FDWW policies. Panel data would offer more robust insights into whether the initial benefits of reduced working hours are sustained, or if new patterns emerge over time, as the current cross-sectional data fails to account for dynamic changes over time and does not follow the same individual over an extended time frame. Furthermore, there is the need for research that can more directly measure the impact of policy interventions such as the FDWW. Experimental and quasi-experimental designs that specifically target the concept of the FDWW, such as randomised controlled trials, with controlled experimental data can provide clearer causal evidence of the effects of reduced working hours on happiness, helping to eliminate biases and reach more definitive conclusions about the causal relationship. Moreover, as the global workforce continues to evolve, researchers could consider examining the potential of reduced working hours and a shift to a FDWW not only in relation to happiness, but also in addressing contemporary issues such as mental health, gender inequalities, labour shortages, work-life integration, and environmental sustainability, which are pressing concerns faced by individuals and societies. Additionally, studying unobserved variables including job satisfaction or social support at work are crucial elements that need to be integrated into future work, as these could potentially explain why some studies - such as this one - fail to find negative effects of working hours on SWB (Booth & van Ours, 2008; Golden & Wiens-Tuers, 2006; Gray et al., 2004; Willson & Dickerson, 2010). Finally, an overview of current issues offers a broader context and insight into the potential of this research question. The global outlook on working hours has been significantly influenced by the COVID-19 pandemic, which has accelerated shifts towards more flexible and remote working arrangements and prompted many to reconsider traditional work models, highlighting the importance of well-being and work-life balance. However, counter-examples such as the six-day workweek in Greece raise the need for further examination of variations across countries to understand the wider implications and provide a holistic understanding of the FDWW with its benefits and drawbacks.

Overall, future research can build on the findings of this work, and given its scale and global reach, the results of this thesis can enhance the current literature and add to the discourse around the FDWW by providing a foundation for further analysis of the relationship between working hours and

happiness, as well as informing the development of an effective and equitable FDWW labour policy.

6. Conclusion

Taking into consideration what has been said so far, we can draw a conclusion about the research question at hand, that is, whether the FDWW has the potential to provide a solution to greater happiness. Due to increasing interest in work-life balance and the potential benefits of reduced working hours, the FDWW has gained popularity for representing a progressive shift in labour policy aimed at improving employee well-being. However, despite fundamental shifts in how society views the balance between work and well-being, rigorous empirical evidence supporting the effectiveness of the FDWW is lacking, making it crucial to investigate both the underlying relationship between working hours and happiness and the effect of the FDWW itself, in order to contribute to a better understanding of how working hours and happiness are related in the context of a FDWW experiment. Therefore, the present thesis aimed at highlighting a segment of the happiness economics literature by empirically evaluating the FDWW experiment in Iceland, thus closing a gap in literature surrounding the discussion about the relationship between working hours and happiness, as well as evaluations of concepts of future of work.

Leveraging a robust dataset from the ESS to scrutinise the association of working hours on employee happiness, with a particular focus on the context of the FDWW in Iceland, this thesis employed a weighted, clustered, and stratified OLS regression analysis to conduct an evidence-based policy evaluation, finding (i) working hours to not be a significant determinant of happiness, and (ii) no evidence to suggest that - qualitatively speaking - the FDWW is a future-proof concept to increase happiness scores. Once employed, there are no significant differences in happiness scores with changes in working hours, suggesting that either working hours are not a substantial determinant of happiness in Iceland, or that there may be other unobserved factors changing with working hours that affect happiness, but are not represented by the model. This aligns with the idea that happiness might be influenced not only by leisure time, but also by the status and financial security conferred by work, supporting concepts such as the Easterlin Paradox and challenging traditional economic theories such as the disutility of labour. The FDWW therefore seems to be rather a product of contemporary societal trends and expectations, fuelled by attention from media, rather than an effective labour policy solution for increasing overall happiness.

These results, while insightful, must be interpreted within the context of several limitations, including (i) the observational nature of the ESS data, meaning that no causal interpretation of the FDWW's impact can be made, (ii) reverse causality causing potential endogeneity of working hours remaining a concern, as individuals may adjust their working hours in response to their levels of satisfaction, and (iii) the

dataset primarily encompassing employees from service sectors, excluding significant economic sectors like manufacturing and fisheries, hence limiting the generalisability of the findings across Iceland's labour market. These limitations are significant, as future studies can increase the validity and credibility of results by considering these drawbacks.

Lastly, this thesis also reveals five research gaps for future studies to explore and analyse more comprehensively. Future research should aim at (i) utilising panel data to better capture the long-term effects of FDWW policies, (ii) providing clearer causal evidence of the effects of reduced working hours on happiness by working with specific FDWW data from experimental and quasi-experimental designs, (iii) considering examining shifts to a FDWW also in relation to other contemporary issues such as mental health, gender inequalities, labour shortages, work-life integration, and environmental sustainability, (iv) studying unobserved variables such as job satisfaction or social support at work in more detail, and (v) examining the FDWW in different societal contexts, raised by counter-examples such as the one from the six-day workweek in Greece, to provide a more holistic understanding of the FDWW.

In summary, there is no evidence to suggest that the FDWW has the potential to provide a solution to greater happiness, as there is no significant association between working hours and happiness and the impact remains unclear. While research into this phenomenon is to be expected to grow at a higher rate than in the previous decade due to evolving dynamics of the workforce, the FDWW is likely to remain a promising trial concept for wealthy countries only. The FDWW symbolises a collective yearning for a future where productivity does not come at the expense of well-being, with this thesis underscoring the complexity of achieving this ideal, revealing that simplistic solutions may fall short of addressing the multifaceted nature of human happiness. Nevertheless, the pursuit of understanding and improving the interplay between working hours and happiness remains a vital endeavour, with the findings of this thesis offering valuable insights for future public and workplace policies, suggesting that a one-size-fits-all approach may not be effective and that labour policies should be tailored to specific national and cultural contexts. Working less and living more may not be achievable with a FDWW as of yet, but it calls upon policymakers, researchers, and society at large to continue exploring, questioning, innovating, and remaining committed to finding pathways that honour both the dignity of work and the sanctity of personal well-being, ultimately creating a more fulfilling and happier life for all.

References

- Alesina, A., Di Tella, R., & MacCulloch, R. (2004). Inequality and Happiness: Are Europeans and Americans Different? *Journal of Public Economics*. <https://doi.org/10.1016/j.jpubeco.2003.07.006>
- Allen, R. E., & Hawes, D. K. (1979). Attitudes toward work, leisure and the four day workweek. *Human Resource Management*, 18(1), 5–10. <https://doi.org/10.1002/hrm.3930180103>

- Altman, M. (2005). *Alternative Economic Approaches to Analyzing Hours of Work Regulation and Reform* (SSRN Scholarly Paper No. 1539785). <https://papers.ssrn.com/abstract=1539785>
- Andrews, F. M., & Withey, S. B. (1976). *Social Indicators of Well-Being: Americans' Perceptions of Life Quality*. Springer. <https://link.springer.com/book/10.1007/978-1-4684-2253-5>
- Anttila, T., Härmä, M., & Oinas, T. (2021). Working hours – tracking the current and future trends. *Industrial Health*, 59(5), 285–292. <https://doi.org/10.2486/indhealth.2021-0086>
- Argyle, M., & Crossland, J. (1987). The dimensions of positive emotions. *British Journal of Social Psychology*, 26(2), 127–137. <https://doi.org/10.1111/j.2044-8309.1987.tb00773.x>
- Arslan, A., & Bektaş, Ç. (2021). Do we work less to be happy? The correlation between happiness and working hours. *Smart Journal*, 7(40), 01–07. <https://doi.org/10.31576/smryj.742>
- Ashta, A. (2000). France Shortens Work-Week: Sharing Scarce Jobs. *Economic and Political Weekly*. <https://doi.org/10.2307/4409780>
- Bakker, A. B., & Demerouti, E. (2007). The Job Demands-Resources model: State of the art. *Journal of Managerial Psychology*, 22(3), 309–328. <https://doi.org/10.1108/02683940710733115>
- Bannai, A., & Tamakoshi, A. (2014). The association between long working hours and health: A systematic review of epidemiological evidence. *Scandinavian Journal of Work, Environment & Health*, 40(1), 5–18. <https://doi.org/10.5271/sjweh.3388>
- Barrick, M. R., & Mount, M. K. (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology*, 44(1), 1–26. <https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>
- Blanchflower, D. G., & Oswald, A. (2011). International Happiness: A New View on the Measure of Performance. *Academy of Management Perspectives*, 25, 6–22. <https://doi.org/10.5465/AMP.2011.59198445>
- Blanchflower, D. G., & Oswald, A. J. (2004). Well-being over time in Britain and the USA. *Journal of Public Economics*, 88(7), 1359–1386. [https://doi.org/10.1016/S0047-2727\(02\)00168-8](https://doi.org/10.1016/S0047-2727(02)00168-8)
- Booth, A. L., & van Ours, J. C. (2008). Job Satisfaction and Family Happiness: The Part-Time Work Puzzle. *The Economic Journal*, 118(526), F77–F99. <https://doi.org/10.1111/j.1468-0297.2007.02117.x>
- Bulut, S., & Maraba, D. (2021). Generation Z and its Perception of Work through Habits, Motivations, Expectations Preferences, and Work Ethics. *Psychology and Psychotherapy: Research Study*, 4(4), 1–5. <https://crimsonpublishers.com/pprs/fulltext/PPRS.000593.php>
- Campbell, A. (1976). Subjective measures of well-being. *American Psychologist*, 31(2), 117–124. <https://doi.org/10.1037/0003-066X.31.2.117>
- Campbell, T. T. (2023). The four-day work week: A chronological, systematic review of the academic literature. *Management Review Quarterly*. <https://doi.org/10.1007/s11301-023-00347-3>
- Caruso, C. C., Bushnell, T., Eggerth, D., Heitmann, A., Kojola, B., Newman, K., Rosa, R. R., Sauter, S. L., & Vila, B. (2006). Long working hours, safety, and health: Toward a National Research Agenda. *American Journal of Industrial Medicine*, 49(11), 930–942. <https://doi.org/10.1002/ajim.20373>
- Cha, Y., & Weeden, K. A. (2014). Overwork and the slow convergence in the gender gap in wages. *American Sociological Review*, 79(3), 457–484. <https://doi.org/10.1177/0003122414528936>
- Chung, H. (2022). A Social Policy Case for a Four-Day Week. *Journal of Social Policy*, 51(3), 551–566. <https://doi.org/10.1017/S0047279422000186>
- Chung, H., & van der Horst, M. (2018). Women's employment patterns after childbirth and the perceived access to and use of flexitime and teleworking. *Human Relations*, 71(1), 47–72. <https://doi.org/10.1177/0018726717713828>
- Clark, A. (2003). Unemployment as a Social Norm: Psychological Evidence from Panel Data. *Journal of Labor Economics*, 21, 289–322. <https://doi.org/10.1086/345560>
- Clark, A., Frijters, P., & Shields, M. (2008). Relative Income, Happiness and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature*, 46, 95–144. <https://doi.org/10.1257/jel.46.1.95>
- Clark, A., & Lelkes, O. (2006). Deliver Us From Evil: Religion as Insurance. https://www.researchgate.net/publication/5081440_Deliver_Us_From_Evil_Religion_as_Insurance
- Coote, A., Franklin, J., & Simms, A. (2010). *21 Hours. Why a Shorter Working Week Can Help Us All to Flourish in the 21st Century*. New Economics Foundation. https://neweconomics.org/uploads/files/f49406d81b9ed9c977_p1m6ibgje.pdf
- Crouter, A. C., Bumpus, M. F., Head, M. R., & McHale, S. M. (2001). Implications of Overwork and Overload for the Quality of Men's Family Relationships. *Journal of Marriage and Family*, 63(2), 404–416. <https://doi.org/10.1111/j.1741-3737.2001.00404.x>
- Deci, E. L., & Ryan, R. M. (2009). The 'What' and 'Why' of Goal Pursuits: Human Needs and the Self-Determination of Behavior: Psychological Inquiry: Vol 11, No 4. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Di Tella, R., & MacCulloch, R. (2006). Some Uses of Happiness Data in Economics. *Journal of Economic Perspectives*, 20(1), 25–46. <https://doi.org/10.1257/089533006776526111>
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, 95(3), 542–575. <https://doi.org/10.1037/0033-2909.95.3.542>
- Diener, E., Lucas, R., & Oishi, S. (2002). Subjective Well-Being: The Science of Happiness and Life Satisfaction. In *Oxford Handbook of Positive Psychology*. <https://doi.org/10.1093/oxfordhb/9780195187243.013.0017>
- Diener, E., Ng, W., Harter, J., & Arora, R. (2010). Wealth and happiness across the world: Material prosperity predicts life evaluation, whereas psychosocial prosperity predicts positive feeling. *Journal of Personality and Social Psychology*, 99(1), 52–61. <https://doi.org/10.1037/a0018066>
- Diener, E., Oishi, S., & Tay, L. (2018). *Handbook of Well-Being*. <https://digitalcommons.unomaha.edu/psychfacbooks/9/>
- Dolan, P., Peasegood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology*, 29(1), 94–122. <https://doi.org/10.1016/j.joep.2007.09.001>
- Dolan, P., & White, M. P. (2007). How Can Measures of Subjective Well-Being Be Used to Inform Public Policy? *Perspectives on Psychological Science*, 2(1), 71–85. <https://doi.org/10.1111/j.1745-6916.2007.00030.x>
- Dong, R., Wu, H., Ni, S., & Lu, T. (2023). The nonlinear consequences of working hours for job satisfaction: The moderating role of job autonomy. *Current Psychology*, 42(14), 11849–11870. <https://doi.org/10.1007/s12144-021-02463-3>
- Easterlin, R. A. (1974). Does Economic Growth Improve the Human Lot? Some Empirical Evidence. In *Nations and Households in Economic Growth* (pp. 89–125). <https://doi.org/10.1016/B978-0-12-205050-3.50008-7>
- Easterlin, R. A. (2005). Building a Better Theory of Well-Being. In L. Bruni & P. L. Porta (Eds.), *Economics and Happiness: Framing the Analysis*. Oxford University Press. <https://doi.org/10.1093/0199286280.003.0002>
- Easterlin, R. A., & O'Connor, K. (2020). *The Easterlin Paradox* (SSRN Scholarly Paper No. 3743147). <https://papers.ssrn.com/abstract=3743147>
- ESS. (2024a). About ESS | European Social Survey. <https://www.europeansocialsurvey.org/about-ess>
- ESS. (2024b). ESS Data Portal. <https://ess.sikt.no/en/?tab=overview>
- ESS. (2024c). Methodology Overview | European Social Survey. <https://www.europeansocialsurvey.org/methodology/methodology-overview>
- ESS. (2024d). Weighting | European Social Survey. <https://www.europeansocialsurvey.org/methodology/ess-methodology/data-processing-and-archiving/weighting>
- Eydal, G., & Gíslason, I. (2014). Family policies: The case of Iceland. In *Handbook of Family Policies across the Globe* (pp. 109–124). https://doi.org/10.1007/978-1-4614-6771-7_8
- Facer, R., & Wadsworth, L. (2008). Alternative Work Schedules and Work-Family Balance: A Research Note. *Review of Public Personnel Administration - REV PUBLIC PERS ADM*, 28, 166–177. <https://doi.org/10.1177/0734371X08315138>

- Félagsmálaráðuneytið. (2019). *Tilraunaverkefni um styttingu vinnuvikunnar hjá ríkinu* (tech. rep.). BSRB. <https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://www.bsrb.is/static/files/Ymis-skjol/skyrsla-rikisins-um-styttingu-vinnuvikunnar-april-2019.pdf&ved=2ahUKEwiO2fPVLsKFAxXoVEEAHQ2UBDcQFnoECBUQAQ&usq=AOvVaw2ILHqmNckU1KYthsVW0cn9>
- Ferrer-i-Carbonell, A., & Frijters, P. (2004). How Important is Methodology for the Estimates of the Determinants of Happiness? *The Economic Journal*, 114(497), 641–659. <https://doi.org/10.1111/j.1468-0297.2004.00235.x>
- Ferrer-i-Carbonell, A., & Gowdy, J. M. (2007). Environmental degradation and happiness. *Ecological Economics*, 60(3), 509–516. <https://doi.org/10.1016/j.ecolecon.2005.12.005>
- Fishburn, P. C. (1968). Utility Theory. *Management Science*, 14(5), 335–378. <https://doi.org/10.1287/mnsc.14.5.335>
- Flouri, E. (2004). Subjective well-being in midlife: The role of involvement and closeness to parents in childhood. *Journal of Happiness Studies: An Interdisciplinary Forum on Subjective Well-Being*, 5(4), 335–358. <https://doi.org/10.1023/B:JOHS.0000048461.21694.92>
- Frey, B. S. (2018). *Economics of Happiness*. Springer. <https://link.springer.com/content/pdf/10.1007/978-3-319-75807-7.pdf>
- Frey, B. S., Benesch, C., & Stutzer, A. (2007). Does watching TV make us happy? *Journal of Economic Psychology*, 28(3), 283–313. <https://doi.org/10.1016/j.joep.2007.02.001>
- Frey, B. S., & Stutzer, A. (2002). What Can Economists Learn From Happiness Research? *J. Econ. Lit.*, 40. <https://doi.org/10.1257/002205102320161320>
- Gallie, D. (2013). *Economic Crisis, Quality of Work, and Social Integration. The European Experience*. Oxford University Press. <https://www.esri.ie/system/files?file=media/file-uploads/2015-07/OPEA95.pdf>
- Gatlin-Keener, C., & Lunsford, R. (2019). Four-Day Workweek: The Microsoft Japan Experience. <https://www.aabri.com/VC2020Manuscripts/VC20032.pdf>
- Golden, L. (2012). *The Effects of Working Time on Productivity and Firm Performance, Research Synthesis Paper* (ILO Conditions of Work and Employment Series No. 33). https://www.researchgate.net/publication/256035039_The_Effects_of_Working_Time_on_Productivity_and_Firm_Performance_Research_Synthesis_Paper
- Golden, L., & Okulicz-Kozaryn, A. (2015). *Work Hours and Worker Happiness in the US: Weekly Hours, Hours Preferences and Schedule Flexibility* (SSRN Scholarly Paper No. 2563374). <https://doi.org/10.2139/ssrn.2563374>
- Golden, L., & Wiens-Tuers, B. (2006). To your happiness? Extra hours of labor supply and worker well-being. *The Journal of Socio-Economics*, 35(2), 382–397. <https://doi.org/10.1016/j.socec.2005.11.039>
- Graham, C. (2005). Insights on Development from the Economics of Happiness. *The World Bank Research Observer*, 20(2), 201–231. <https://doi.org/10.1093/wbro/lki010>
- Graham, C., Eggers, A., & Sukhtankar, S. (2004). Does Happiness Pay? An Exploration Based on Panel Data from Russia. *Journal of Economic Behavior & Organization*, 55, 319–342. <https://doi.org/10.1016/j.jebo.2003.09.002>
- Gray, M., Qu, L., Stanton, D., & Weston, R. (2004). Long Work Hours and the Wellbeing of Fathers and Families. *Australian Journal of Labour Economics*, 7, 255–273. <https://ideas.repec.org/p/wpa/wuwpla/0405007.html>
- Gudmundsdóttir, D. G. (2013). The Impact of Economic Crisis on Happiness. *Social Indicators Research*, 110(3), 1083–1101. <https://doi.org/10.1007/s11205-011-9973-8>
- Gudmundsdóttir, D. G., Ásgeirsdóttir, B. B., Huppert, F. A., Sigfúsdóttir, I. D., Valdimarsdóttir, U. A., & Hauksdóttir, A. (2016). How Does the Economic Crisis Influence Adolescents' Happiness? Population-Based Surveys in Iceland in 2000–2010. *Journal of Happiness Studies*, 17(3), 1219–1234. <https://doi.org/10.1007/s10902-015-9639-3>
- Gunderson, R. (2018). Work time reduction and economic democracy as climate change mitigation strategies: Or why the climate needs a renewed labor movement. *Journal of Environmental Studies and Sciences*, 9. <https://doi.org/10.1007/s13412-018-0507-4>
- Haraldsson, G., & Kellam, J. (2021). *Going Public: Iceland's journey to a shorter working week* (tech. rep.). Alda and Autonomy. https://autonomy.work/wp-content/uploads/2021/06/ICELAND_4DW.pdf
- Hayden, A. (2006). France's 35-Hour Week: Attack on Business? Win-Win Reform? Or Betrayal of Disadvantaged Workers? *Politics & Society*, 34(4), 503–542. <https://doi.org/10.1177/0032329206293645>
- Hayes, M., Chumney, F., & Buckingham, M. (2022). *Global Workplace Study 2022* (tech. rep.). <https://www.adpri.org/assets/global-workplace-study-2022/>
- Helliwell, J. F. (2003). How's life? Combining individual and national variables to explain subjective well-being. *Economic Modelling*, 20(2), 331–360. [https://doi.org/10.1016/S0264-9993\(02\)00057-3](https://doi.org/10.1016/S0264-9993(02)00057-3)
- Howell, R. T., & Howell, C. J. (2008). The relation of economic status to subjective well-being in developing countries: A meta-analysis. *Psychological Bulletin*, 134(4), 536–560. <https://doi.org/10.1037/0033-2909.134.4.536>
- Johnson, S., Li, J., Kendall, G., Strazdins, L., & Jacoby, P. (2013). Mothers' and Fathers' Work Hours, Child Gender, and Behavior in Middle Childhood. *Journal of Marriage and Family*, 75(1), 56–74. <https://doi.org/10.1111/j.1741-3737.2012.01030.x>
- Kahneman, D., & Deaton, A. (2010). High Income Improves Evaluation of Life But Not Emotional Well-Being. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 16489–16493. <https://doi.org/10.1073/pnas.1011492107>
- Kallis, G., Kalush, M., Flynn, H. O., Rossiter, J., & Ashford, N. (2013). "Friday off": Reducing Working Hours in Europe. *Sustainability*, 5(4), 1545–1567. <https://doi.org/10.3390/su5041545>
- Kameråde, D., Wang, S., Burchell, B., Balderson, S. U., & Coutts, A. (2019). A shorter working week for everyone: How much paid work is needed for mental health and well-being? *Social Science & Medicine*, 241, 112353. <https://doi.org/10.1016/j.socscimed.2019.06.006>
- Kaminska, O., & Lynn, P. (2017). Survey-Based Cross-Country Comparisons Where Countries Vary in Sample Design: Issues and Solutions. *Journal of Official Statistics*, 33(1), 123–136. <https://doi.org/10.1515/jos-2017-0007>
- Kezer, M., & Cemalcilar, Z. (2020). A Comprehensive Investigation of Associations of Objective and Subjective Socioeconomic Status with Perceived Health and Subjective Well-Being. *ADRIPS*, 33(1), Article 1. <https://doi.org/10.5334/irsp.364>
- Killingsworth, M. (2021). Experienced well-being rises with income, even above \$75,000 per year. *Proceedings of the National Academy of Sciences*, 118, e2016976118. <https://doi.org/10.1073/pnas.2016976118>
- Kleiner, S., & Pavalko, E. K. (2010). Clocking In: The Organization of Work Time and Health in the United States. *Social Forces*, 88(3), 1463–1486. <https://doi.org/10.1353/sof.0.0301>
- Kobie, N. (2021). What really happened in Iceland's four-day week trial. *Wired*. <https://www.wired.com/story/iceland-four-day-work-week/>
- Krueger, A. B. (Ed.). (2009). *Measuring the Subjective Well-Being of Nations: National Accounts of Time Use and Well-Being*. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/M/bo6683637.html>
- Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. (2005). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science (New York, N.Y.)*, 306, 1776–1780. <https://doi.org/10.1126/science.1103572>
- Kuhn, U. (2019). *Measurement of income in surveys | FORS* (FORS Guide No. 02 (Version 1.0)). Lausanne: Swiss Centre of Expertise in the Social Sciences FORS. <https://doi.org/10.24449/FG-2019-00002>
- Kunn-Nelen, A., Grip, A., & Fouarge, D. (2013). Is Part-Time Employment Beneficial for Firm Productivity? *Industrial and Labor Relations Review*, 66, 1172–1191. <https://doi.org/10.1177/001979391306600507>
- Laurence, G. A., Fried, Y., & Raub, S. (2016). Evidence for the need to distinguish between self-initiated and organizationally imposed over-

- load in studies of work stress. *Work & Stress*, 30(4), 337–355. <https://doi.org/10.1080/02678373.2016.1253045>
- Liu, B., Chen, H., & Gan, X. (2019). How Much Is Too Much? The Influence of Work Hours on Social Development: An Empirical Analysis for OECD Countries. *International Journal of Environmental Research and Public Health*, 16(24), 4914. <https://doi.org/10.3390/ijerph16244914>
- Liu, B., Chen, H., & Huang, X. (2018). Map Changes and Theme Evolution in Work Hours: A Co-Word Analysis. *International Journal of Environmental Research and Public Health*, 15(5), Article 5. <https://doi.org/10.3390/ijerph15051039>
- Liu, B., Floud, S., Pirie, K., Green, J., Peto, R., & Beral, V. (2015). Does happiness itself directly affect mortality? The prospective UK Million Women Study. *The Lancet*, 387. [https://doi.org/10.1016/S0140-6736\(15\)01087-9](https://doi.org/10.1016/S0140-6736(15)01087-9)
- Louis, V. V., & Zhao, S. (2002). Effects of Family Structure, Family SES, and Adulthood Experiences on Life Satisfaction. *Journal of Family Issues*, 23(8), 986–1005. <https://doi.org/10.1177/019251302237300>
- Luhmann, M., Hawley, L. C., Eid, M., & Cacioppo, J. T. (2012). Time frames and the distinction between affective and cognitive well-being. *Journal of Research in Personality*, 46(4), 431–441. <https://doi.org/10.1016/j.jrp.2012.04.004>
- Luhmann, M., Hofmann, W., Eid, M., & Lucas, R. E. (2012). Subjective well-being and adaptation to life events: A meta-analysis. *Journal of Personality and Social Psychology*, 102(3), 592–615. <https://doi.org/10.1037/a0025948>
- Luttmer, E. F. P. (2005). Neighbors as Negatives: Relative Earnings and Well-Being*. *The Quarterly Journal of Economics*, 120(3), 963–1002. <https://doi.org/10.1093/qje/120.3.963>
- MacKerron, G. (2012). Happiness Economics from 35 000 Feet. *Journal of Economic Surveys*, 26(4), 705–735. <https://doi.org/10.1111/j.1467-6419.2010.00672.x>
- Mahoney, T. A. (1978). The Rearranged Work Week: Evaluations of Different Work Schedules. *California Management Review*, 20(4), 31–39. <https://doi.org/10.2307/41164779>
- Mahoney, T. A., Newman, J. M., & Frost, P. J. (1975). Workers' Perceptions of the Four-Day Week. *California Management Review*, 18(1), 31–35. <https://doi.org/10.2307/41164623>
- McMahon, D. M. (2008). The pursuit of happiness in history. In *The science of subjective well-being* (pp. 80–93). The Guilford Press. <https://psycnet.apa.org/record/2008-00541-004>
- Meier, S., & Stutzer, A. (2004). *Is Volunteering Rewarding in Itself?* (SSRN Scholarly Paper No. 505042). <https://doi.org/10.2139/ssrn.505042>
- Moore, J. E. (1990). A meta-analysis of the effect of compressed work weeks. *Applied HRM Research*, 1(1), 8–14. [http://applyhrm.as.p.radford.edu/1990/Moore-1\(1\).pdf](http://applyhrm.as.p.radford.edu/1990/Moore-1(1).pdf)
- Muffels, R., & Kemperman, B. (2011). Does a Better Job Match Make Women Happier? *Work Orientations, Work-Care Choices and Subjective Well-Being in Germany* (SOEPpapers on Multidisciplinary Panel Data Research No. 361). DIW Berlin, The German Socio-Economic Panel (SOEP). https://ideas.repec.org/p/diw/diwsop/diw_sp361.html
- Nässén, J., & Larsson, J. (2015). Would shorter working time reduce greenhouse gas emissions? An analysis of time use and consumption in Swedish households. *Environment and Planning C: Government and Policy*, 33(4), 726–745. <https://doi.org/10.1068/c12239>
- Nikolova, M., & Graham, C. (2020). The Economics of Happiness. *GLO Discussion Paper Series*, Article 640. <https://ideas.repec.org/p/zbw/glodps/640.html>
- Norrish, J. M., & Vella-Brodick, D. A. (2008). Is the study of happiness a worthy scientific pursuit? *Social Indicators Research*, 87(3), 393–407. <https://doi.org/10.1007/s11205-007-9147-x>
- OECD. (2009). Adjusting household incomes: Equivalence scales. <https://www.oecd.org/els/soc/OECD-Note-EquivalenceScales.pdf>
- Oesch, D. (2006). Coming to Grips with a Changing Class Structure: An Analysis of Employment Stratification in Britain, Germany, Sweden and Switzerland. *International Sociology*, 21(2), 263–288. <https://doi.org/10.1177/0268580906061379>
- oikonomia. (2024). New Six-Day work week to begin July 1st in Greece. <https://sofokleous10.gr/2024/06/19/new-six-day-work-week-to-begin-july-1st-in-greece/>
- Pencavel, J. (2015). The Productivity of Working Hours. *The Economic Journal*, 125(589), 2052–2076. <https://doi.org/10.1111/eoj.12166>
- Pouwels, B., Siegers, J., & Vlasblom, J. D. (2008). Income, working hours, and happiness. *Economics Letters*, 99, 72–74. <https://doi.org/10.1016/j.econlet.2007.05.032>
- Pullinger, M. (2014). Working time reduction policy in a sustainable economy: Criteria and options for its design. *Ecological Economics*, 103, 11–19. <https://doi.org/10.1016/j.ecolecon.2014.04.009>
- Rätzel, S. (2012). Labour Supply, Life Satisfaction, and the (Dis)Utility of Work*. *The Scandinavian Journal of Economics*, 114(4), 1160–1181. <https://doi.org/10.1111/j.1467-9442.2012.01717.x>
- Rothbard, N. P., & Edwards, J. R. (2003). Investment in Work and Family Roles: A Test of Identity and Utilitarian Motives. *Personnel Psychology*, 56(3), 699–729. <https://doi.org/10.1111/j.1744-6570.2003.tb00755.x>
- Rudolf, R. (2013). Work Shorter, Be Happier? Longitudinal Evidence from the Korean Five-Day Working Policy. *Journal of Happiness Studies*, 15, 1139–1163. <https://doi.org/10.1007/s10902-013-9468-1>
- Schor, J. (2015). Work sharing. In G. D'Alisa, F. Demaria, & G. Kallis (Eds.), *DEGROWTH: A Vocabulary for a New Era* (pp. 195–197). Routledge.
- Shao, Q. (2022). Does less working time improve life satisfaction? Evidence from European Social Survey. *Health Economics Review*, 12(1), 50. <https://doi.org/10.1186/s13561-022-00396-6>
- Smith, S. J. (1986). The Growing Diversity of Work Schedules. *Monthly Labor Review*, 109(11). https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://www.bls.gov/opub/mlr/1986/11/art2full.pdf&ved=2ahUKewjhOnS04CGAxVmVvEDHVSGC50QFnoECBIQAQ&usq=AOvVaw20hbvw-YMvC9jofT0H_kiGB
- Sonnentag, S. (2012). Psychological Detachment From Work During Leisure Time: The Benefits of Mentally Disengaging From Work. *Current Directions in Psychological Science*, 21(2), 114–118. <https://doi.org/10.1177/09637214114134979>
- Stevenson, B., & Wolfers, J. (2013). Subjective Well-Being and Income: Is There Any Evidence of Satiation? *The American Economic Review*, 103(3), 598–604. <https://www.aeaweb.org/articles?id=10.1257/aer.103.3.598>
- Stier, H., & Lewin-Epstein, N. (2003). Time to Work: A Comparative Analysis of Preferences for Working Hours. *Work and Occupations*, 30(3), 302–326. <https://doi.org/10.1177/0730888403253897>
- Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress* (tech. rep.). <https://ec.europa.eu/eurostat/documents/8131721/8131772/Stiglitz-Sen-Fitoussi-Commission-report.pdf>
- Stone, J. (2021). World's largest ever four day week trial in Iceland 'overwhelming success'. *The Independent*. <https://www.independent.co.uk/news/uk/politics/four-day-week-pilot-iceland-b1877171.html>
- Strazdins, L., Baxter, J., & Li, J. (2017). Long Hours and Longings: Australian Children's Views of Fathers' Work and Family Time: Children's Views of Fathers' Work and Family Time. *Journal of Marriage and Family*, 79. <https://doi.org/10.1111/jomf.12400>
- Stutzer, A. (2004). The role of income aspirations in individual happiness. *Journal of Economic Behavior & Organization*, 54(1), 89–109. <https://doi.org/10.1016/j.jebo.2003.04.003>
- Thoits, P. A., & Hewitt, L. N. (2001). Volunteer Work and Well-Being. *Journal of Health and Social Behavior*, 42(2), 115–131. <https://doi.org/10.2307/3090173>
- Uchida, Y., & Rappleye, J. (2023). Happiness: A World Map. In *An Interdependent Approach to Happiness and Well-Being* (pp. 19–33). https://doi.org/10.1007/978-3-031-26260-9_2
- Veenhoven, R. (2008). Sociological theories of subjective well-being. *The Science of Subjective Well-Being: A Tribute to Ed Diener*. <https://repub.eur.nl/pub/14879/2008c-full.pdf>
- Viner, J. (1925). The Utility Concept in Value Theory and Its Critics. *Journal of Political Economy*, 33(4), 369–387. <https://doi.org/10.1086/253690>

- Vlasblom, J. D., & Schippers, J. (2006). Changing dynamics in female employment around childbirth: Evidence from Germany, the Netherlands and the UK. *Work, Employment and Society*, 20(2), 329–347. <https://doi.org/10.1177/0950017006064117>
- VR stéttarfélag. (n.d.). Kjaramál. Know your rights | Employee Relationship. Retrieved May 17, 2024, from <https://www.vr.is/kjaramal/thejektu-thinn-rett/know-your-rights/>
- Willson, J., & Dickerson, A. (2010). *Part time employment and happiness: A cross-country analysis* (Working Paper No. 2010021). The University of Sheffield, Department of Economics. http://www.shef.ac.uk/economics/research/serps/articles/2010_021.html
- Wong, K., Chan, A. H. S., & Ngan, S. C. (2019). The Effect of Long Working Hours and Overtime on Occupational Health: A Meta-Analysis of Evidence from 1998 to 2018. *International Journal of Environmental Research and Public Health*, 16(12), 2102. <https://doi.org/10.3390/ijerph16122102>
- Yamashita, T., Bardo, A. R., & Liu, D. (2016). Are East Asians happy to work more or less? Associations between working hours, relative income and happiness in China, Japan, South Korea and Taiwan. *Asian Journal of Social Psychology*, 19(3), 264–274. <https://doi.org/10.1111/ajsp.12140>
- Young, C. (2012). Losing a Job: The Nonpecuniary Cost of Unemployment in the United States. *Social Forces*, 91(2), 609–633. https://www.researchgate.net/publication/260139851_Losing_a_Job_The_Nonpecuniary_Cost_of_Unemployment_in_the_United_States



Integrating Sustainability in Risk Management and Internal Control Systems: An Empirical Assessment of ESG Reporting of German DAX40 Firms

Niklas Janßen

Leuphana University Lüneburg

Abstract

Integrating sustainability into enterprise risk management (ERM) and internal control system (ICS) and the corresponding reporting is gaining importance for organizations due to increasing pressure from regulators and stakeholders. The objective of this paper is to analyze corporate disclosures on the integration of sustainability into the ERM and the ICS. Based on stakeholder agency theory, this study applies a self-developed disclosure index on the ESG reports of 29 German DAX40 companies for 2022 and 2023. The index evaluates from an organizational, regulatory, and, with a focus on environmental issues, business practice perspective. Overall, companies report with restrained quality. The study found varying results across the index sections and a minor positive trend from 2022 to 2023. A sub analysis to identify potential differences in reporting behavior according to the affiliation to environmentally sensitive sectors revealed mixed findings. This paper holds different methodological limitations. However, it contributes to scarce qualitative research and provides deep insights into corporate sustainability reporting of German DAX40 companies. The results are of practical importance for businesses and regulatory bodies, as they reveal weaknesses in corporate reporting.

Keywords: disclosure index; ESG reporting; internal control system; risk management; sustainable corporate governance

1. Introduction

The 2008/09 financial crisis led to a significant loss of stakeholder trust in corporate governance mechanisms (Earle, 2009), and exposed massive failures in the area of enterprise risk management (ERM) and internal control systems (ICS). While these systems traditionally focused on financial aspects, the integration of sustainability is becoming increasingly important from multiple perspectives. Building on the following considerations, this study focuses on the reporting about the integration of sustainability into the ERM and ICS.

In terms of terminology, this paper defines the ERM as a system for the identification, assessment, reaction and prevention of negative effects (risks) that can significantly impact an organization (Bartuschka, 2022; Boiral et al., 2020; Brown et al., 2009; COSO, 2017; Gatzert & Martin, 2015). The ICS aims to ensure efficiency and effectiveness of operations, as well as compliance with regulations and guidelines

(Bartuschka, 2022; COSO, 2013, 2023; Deumes & Knechel, 2008).

Overarching relevance for the research objective stems from the growing number of sustainability issues and associated risks that organizations face, which can cover a wide range of aspects (Boiral et al., 2020). These issues and risks are difficult to determine, quantify or mitigate, and can affect the profitability, success and ultimately even the survival of a company (COSO and wbcSD, 2018; Crawford & Nilsson, 2023). At the same time, stakeholders increasingly expect reliable and transparent sustainability reporting alongside sustainable corporate performance (Boiral et al., 2020; Fernandez-Feijoo et al., 2014; Rezaee, 2016).

Practical relevance was emphasized in previous academic literature (e.g. D. R. Anderson and Anderson, 2009; Shad et al., 2019). COSO and wbcSD (2018) highlight the integration of sustainability into the ERM and a corresponding reporting, as it helps to identify and respond to sustainability issues at an early stage and creates more transparency to stakehold-

© The Author(s) 2025. Published by Junior Management Science.

This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.

ers on how the organization deals with risks. Moreover, it can enable companies not only to react, but to proactively improve their sustainability performance (Rezaee, 2016). Although companies are already focusing on sustainability and risk management, there is often a lack of integration between these two paradigms (Soomro & Lai, 2017). On the other hand, an effective ICS can provide accountability in organizations (Jones, 2008), or help to assure the accuracy of environmental information collection and processing (Chan et al., 2021; J. Zhang et al., 2024). It therefore enhances the reliability and quality of sustainability disclosures, as credible reporting requires timely, complete and precise information (Healy & Palepu, 2001; Mercer, 2004) and is dependent on the internal collection of information (Traxler et al., 2020). A strong ICS can also enhance the effectiveness of the ERM (COSO and wbcSD, 2018). Integrating sustainability into the systems can therefore be an approach to adequately address sustainability risks and increase the reliability of sustainability reports (Chan et al., 2021; Harasheh & Provasi, 2023; J. Zhang et al., 2024).

Regulatory relevance can be derived from recent developments. Since the financial year 2017, European public-interest entities (PIEs) must publish a non-financial declaration under the 2014 European Union (EU) Non-Financial Reporting Directive (NFRD). Building on the NFRD, the 2022 EU Corporate Sustainability Reporting Directive (CSRD) and related Environmental Sustainability Reporting Standards (ESRS) refer for the first time to the RMS and ICS. PIEs must report on risk management and internal controls over the sustainability reporting process, as well as on selected risk management processes starting from financial year 2024. Future European directives (e.g. Corporate Sustainability Due Diligence Directive (CSDDD)) will add further obligations. In the context of German national legislation, the 2021 Supply Chain Due Diligence Act (SCDDA), effective from 2023, addresses the management of sustainability risks in the supply chain. The revised German Corporate Governance Code (GCGC) of 2022 introduced a soft law recommendation stating that the RMS and ICS should also include sustainability-related objectives.

There is growing research activity addressing ERM and ICS in the context of sustainability. For example, prior studies focused on the practical utilization of ERM from a management perspective (e.g. Valinejad and Rahmani, 2018), or on the implementation of environmental management systems as an approach for risk reduction (e.g. Boiral et al., 2018). Most empirical research uses quantitative methods and indicate positive effects of the ERM and ICS on corporate social responsibility (CSR) performance and disclosure (e.g. P. Huang et al., 2022; Musallam, 2018; Pérez-Cornejo and de Quevedo-Puente, 2023). There are also studies to the contrary (e.g. Sarkis, 2006). However, little qualitative research exists, especially with a focus on the German capital market. For example, a number of authors applied surveys to examine the integration of sustainability into ERM and ICS (DRSC, 2023; Scheffler & Flath, 2023). Other research already analyzed the reports of DAX companies at the intersec-

tion of sustainability and both systems (Stakeholder Reporting, 2022, 2023; Teucher & Ratzinger-Sakel, 2024). However, these studies only addressed sustainability as a subordinate focus or lacked methodological quality. Consequently, a research gap can be identified in the area of a qualitative analysis of corporate reports in the German capital market.

Given the illustrated relevance, this paper closes the identified research gap and contributes to existing literature with the following two research objectives. First, this study aims to provide empirical insights on the reporting quality about sustainability-related ERM and ICS. Second, it builds on previous findings on the reporting behavior of companies operating in (non-)environmental sensitive sectors and develops them in the context of sustainability-related ERM and ICS. Environmentally sensitive sectors (chemicals, construction, industrials, utilities) are defined following Garcia et al. (2017) and Richardson and Welker (2001).

This paper assesses the reporting on the sustainability-related ERM and ICS within the environmental, social and governance (ESG) disclosures of German DAX40 companies for the years 2022 and 2023 from the perspective of stakeholder agency theory. A self-developed disclosure index was applied, which allows an integrated analysis of the reporting on both systems. This method has been used in previous research (e.g. Hooghiemstra et al., 2015; Michelon et al., 2015), although the integrated analysis represents a new approach.

The sample is motivated as follows. DAX40 represent the largest listed firms in the German capital market. From a theoretical perspective, large companies often face significant agency problems (Jensen & Meckling, 1976; Sarens & Christopher, 2010). ERM or ICS can thereby serve as monitoring mechanisms (Hernández-Madrigal et al., 2020). As already mentioned, stakeholders increasingly put pressure on organizations to operate sustainably and demand transparent corporate disclosure. Integrating sustainability into the ERM and ICS, as well as reporting on it, may therefore reduce information asymmetries and agency conflicts. Moreover, the German capital market represents an interesting field of research, as it is the subject of globally unique European provisions (Velte, 2023).

After this introduction, Chapter 2 develops a basic understanding of the systems, and presents the theoretical framework and regulatory environment. Chapter 3 provides an overview of previous literature and develops the research questions. This is followed by a description of the research methodology in Chapter 4. Then, Chapter 5 elaborates on the empirical results and discusses them in light of the theoretical framework. Implications as well as limitations and recommendations for future research can be found in Chapter 6. Finally, Chapter 7 gives a summary.

2. Theoretical framework

2.1. Enterprise risk management and internal control system

There are various definitions of ERM and ICS. This paper adopts the understanding of the widely used ERM frame-

work by the Committee of Sponsoring Organizations of the Treadway Commission (COSO). COSO defines ERM as follows: “The culture, capabilities, and practices, integrated with strategy-setting and performance, that organizations rely on to manage risk in creating, preserving, and realizing value.” (COSO, 2017, p. 10). In this sense and in line with the regulatory environment of this paper, the aim of the ERM is the identification, assessment, reaction and prevention of negative effects (risks) that can significantly impact an organization (Bartuschka, 2022; Boiral et al., 2020; Brown et al., 2009; COSO, 2017; Gatzert & Martin, 2015). Given the aforementioned scope and complexity of sustainability issues, this paper adapts the concept of an enterprise-wide risk management system (RMS), which differs from traditional silo-based RMS through its integration into the entire company (Hernández-Madrigal et al., 2020; Lu et al., 2022; Naseem et al., 2020). For the ICS, the paper also refers to the corresponding COSO framework, which defines the system as follows: “Internal control is a process, effected by an entity’s [...] management, [...], designed to provide reasonable assurance regarding the achievement of objectives relating to operations, reporting, and compliance.” (COSO, 2013, p. 3). Consequently, it aims ensure the efficiency and effectiveness of operations as well as compliance with regulations and guidelines (Bartuschka, 2022; COSO, 2013, 2023; Deumes & Knechel, 2008). This paper adopts the view that the ICS is an essential part of a broader ERM (COSO, 2023; Velte, 2022). This perspective is supported by previous studies and legislation that often refer to internal control and risk management systems and thereby imply closely interconnection (Henk, 2020; Sarens et al., 2009). The systems are nevertheless regarded as separate.

Against the background of the geographically applicable regulations in this paper, an effective risk management and ICS requires internal auditing and includes a compliance management system (CMS) (Government Commission on the German Corporate Governance Code, 2022). The internal audit system (IAS) provides auditing and consulting services on behalf of the management, with the aim of evaluating and improving the effectiveness of the ERM and ICS (Berwanger & Kullmann, 2012), while the CMS aims to ensure compliance with regulatory and legal obligations (IDW, 2022). The Institute of German Certified Public Accountants (IDW) has established auditing standards for all four systems (RMS, ICS, CMS, IAS) (see IDW, 2017a, 2017b, 2017c, 2022). It can therefore be assumed that the systems are closely linked but can be observed separately. Given the focus of this paper, the CMS and IAS are not considered in detail in the remainder of this paper.

2.2. Stakeholder agency theory

This paper will examine ESG reports on the sustainability-related ERM and ICS from the perspective of stakeholder agency theory (Hill & Jones, 1992). This theory represents a trade-off between stakeholder theory (Freeman, 1984; see also Freeman, 2010) and agency theory (Hill & Jones,

1992; Jensen & Meckling, 1976). According to agency theory, information asymmetries arise between management and shareholders due to the separation of ownership and control (Jensen & Meckling, 1976). This can lead to self-serving behavior and moral hazard by the management (Brennan, 2005). Accordingly, suitable control mechanisms for monitoring the management must be established as part of corporate governance (Short et al., 1999). Hill and Jones (1992) extended this problem of agency conflicts between the management and shareholders to other stakeholders (Hernández-Madrigal et al., 2020). Based on this view, the company is a nexus of contracts with explicit and implicit contractual relationships involving various stakeholders, while simultaneously recognizing different levels of power (Collier, 2008). Therefore, managers can be considered as stakeholder agents (Naciti et al., 2021). Stakeholders include for example investors, customers, employees, the environment, or the society (Hill & Jones, 1992; Rezaee, 2016).

The objective of this research is to analyze the reporting practices on the integration of sustainability into the ERM and ICS of German DAX 40 companies that were initiated by the management. Unlike the Anglo-Saxon one-tier system, the German listed companies are organized under the so-called two-tier system (management and supervisory board) (Stiglbauer & Velte, 2014). While this organizational separation can also lead to information asymmetries within an organization, the theoretical framework of this paper focuses on the relationship between management and external stakeholders.

Stakeholder agency theory may provide various contributions to the different facets of this research. Stakeholders are expecting disclosure about the organization’s activities and how it addresses risks in order to ensure the company’s survival (Shad et al., 2019). As a result of the growing focus on sustainability issues, companies publish ESG reports with the aim of increasing transparency and reducing information asymmetries (Buallay, 2022; Gray et al., 1995; Hahn & Kühnen, 2013; Michelon & Parbonetti, 2012). Management may see an increased need for the publication of sustainability information if the company is undervalued on the capital market, for example (Dienes & Velte, 2016). Additional incentives for publishing sustainability reports come from standard setters such as the Global Reporting Initiative (GRI) (Dienes & Velte, 2016). This framework is used as the dominant reporting basis for the sustainability reports of German DAX40 companies (Stakeholder Reporting, 2023), thereby supporting the theoretical focus of this paper.

DAX40 companies represent the largest listed companies within the German capital market. Following Dumay and Hossain (2019), larger companies are likely to publish more information to lower information asymmetry and reduce agency costs, as they are often facing significant agency problems (Jensen & Meckling, 1976; Sarens & Christopher, 2010).

The substantive focus of this paper lies on the reporting on the sustainability-related ERM and ICS. In general, both

systems are not directly visible to stakeholders outside the organization, which makes them dependent on reporting in light of information asymmetries (Deumes & Knechel, 2008). However, reporting on the sustainability-related risk management and control systems, as well as the integration of sustainability itself, is not directly prescribed in Germany yet. But, voluntary disclosure may serve as an external monitoring mechanism to reduce agency costs (Louie et al., 2019). Stakeholder agency theory could therefore explain why companies voluntarily report on their sustainability-related ERM and ICS. Besides, this paper deals with the integration of sustainability into ERM and ICS. As described in the introduction, ERM and ICS can provide several benefits, such as enhancing environmental information or help to adequately address sustainability risks. Therefore, a corresponding risk management and internal control system (including internal auditing) can help to reduce the efficiency loss in agency conflicts by serving as a monitoring and control mechanism according to stakeholders' demands (D. Anderson et al., 1993; Deumes & Knechel, 2008; Hernández-Madrigal et al., 2020; Jensen & Meckling, 1976; Sarens & Christopher, 2010).

A recent literature review on sustainability reporting and management control showed that stakeholder theory and agency theory are the most commonly used theories in this area (Traxler et al., 2020). When analyzing the ICS, agency theory is also the most popular theory, as shown by the literature review of Henk (2020). The same applies to ERM (Crawford & Jabbour, 2024; Jankensgård, 2019). Stakeholder agency theory is used less frequently, but has also been applied by previous studies, for example in the analysis between ERM and CSR (e.g. Pérez-Cornejo and de Quevedo-Puente, 2023) or ERM implementation (e.g. Hernández-Madrigal et al., 2020).

2.3. Regulatory background

Given the overall research objective, there are several regulatory dimensions of relevance. German DAX40 companies have been subject to national legislation with direct reference to the RMS and ICS for a long time. The obligation to implement a RMS was introduced for the first time in 1999 with the German law on control and transparency in business (KonTraG), thereby laying the foundation for further legal obligations. In the following decades, often in response to economic scandals, a series of laws followed, which placed more specific requirements on the RMS and ICS and its reporting (e.g. the German Accounting Law Modernization Act (BilMoG), or the German Financial Market Integrity Strengthening Act (FISG)). Sustainability was first considered with the recommendations of the revised version of the GCGC effective from fiscal year 2022.

Next to national legislation, German DAX40 companies are also subject to European regulatory requirements, which play a unique role in global comparison (Velte, 2023). General relevance for the integration of sustainability into the RMS and ICS can be derived from the requirements of the European Green Deal. According to its provisions, European

companies must be climate-neutral by 2050 (European Commission, 2021). In the context of this paper, this target is relevant for companies as the RMS and ICS can also focus on the reduction of emissions (COSO and wbcscd, 2018), for example by the integration of carbon risks in the RMS (Subramaniam et al., 2015), or by enhancing quality of greenhouse gas data through an effective ICS (COSO, 2023). More specific requirements for the integration of sustainability into the RMS and ICS and the corresponding reporting are set out in the CSRD and the associated ESRS, which are to be applied from the 2024 reporting year.

The described regulatory developments at national and European level underline the complexity of the regulatory environment for the German capital market and DAX40 companies. Due to the large number of individual regulations both at national and European level, Table B1 in the Appendix provides a comprehensive overview of the relevant standards regarding the RMS and ICS and sustainability reporting.

3. Literature review and research questions

There is already considerable research addressing the integration of sustainability into ERM and ICS from a conceptual, management perspective. For example, studies look at ERM in relation to different areas of application, such as the supply chain (e.g. Foerstl et al., 2010; Giannakis and Papadopoulos, 2016; Valinejad and Rahmani, 2018; T. Wu and Blackhurst, 2009) or sustainable risk identification in the product development (e.g. Palousis et al., 2010). Lenssen et al. (2014), for example, emphasizes the relevance of an integrative approach for sustainability, risk management and governance. Other studies highlight the advantages of an ICS in relation to sustainability issues in organizations, for example with regard to corporate sustainability performance (e.g. Harasheh and Provasi, 2023), or data collection (e.g. Chan et al., 2021; J. Zhang et al., 2024).

Empirically, the link between ERM and ICS and sustainability has also already been investigated. With regard to ERM, there is a large number of studies that examine the influence of environmental management systems (EMS). Mohammed (2000) found that one of the main reasons for implementing an EMS can be the reduction of risks. Therefore it can be argued that EMS is a crucial element of a complete ERM system (S. J. Wu et al., 2008). The results of a literature review and a meta-analysis point in the same direction. Boiral et al. (2018) found mainly positive effects on the reduction of environmental risks in the application of the EMS Standard ISO 14001. The results of Nawrocka and Parker (2009) show similar results, but with less available evidence. As an EMS can also include other factors in addition to the management of risks, it will not be discussed in detail in the remainder of this paper.

Most of prior research showed a positive effect of RMS on CSR applying quantitative methods. Pérez-Cornejo and de Quevedo-Puente (2023) found that ERM system quality can improve CSR performance and corporate reputation, while

Musallam (2018) showed that the existence of a RMS enhances CSR disclosure. Moreover, firms with an effective RMS can reduce the effects of individual and overarching risks on CSR objectives, and are more willing to engage in CSR activities (Kuo et al., 2021). In line with that, Shah et al. (2024) demonstrated that the integration of an ERM reduces ESG risks and increases “green growth”. On the other hand, there are limited studies that indicate a negative effects. Sarkis (2006) showed a significant negative effect of adaption of environmental risk management practices on environmental performance, with a potential explanation that the focus on risk minimization does not necessarily entail environmental improvements. In line with that, Dobler et al. (2014) found a negative but insignificant relationship between active management of environmental risks and environmental performance. Here the authors argued that this finding does not necessarily indicate an ineffective RMS, but that the system simply did not affect the examined environmental performance variables.

In other studies, ERM, CSR or ESG was applied as a moderator or mediator. ERM mediates the positive correlation between board gender diversity and sustainability performance (Fakir & Jusoh, 2020), or partially mediates the positive relationship between CSR and firm performance (Naseem et al., 2020). Chairani and Siregar (2021) found that ESG moderates the positive relationship between ERM and firm value, while Pérez-Cornejo and de Quevedo-Puente (2023) showed that CSR mediates the positive correlation between ERM and corporate reputation. Further studies indicated additional effects of integrating sustainability into ERM, such as the reduction of costs of equity capital (Sharfman & Fernando, 2008), or the financial risk forecasting effectiveness (Capelli et al., 2021). Another widely cited work is that of Shad et al. (2019), who establishes a conceptual framework for the integration of sustainability reporting into ERM and the relationship to business performance.

For the ICS, recent studies empirically indicate a positive influence of the ICS on CSR disclosure or performance. In general, ethical and social responsible firms are more likely to have an effective ICS (Kim et al., 2017). P Huang et al. (2022) found in a Chinese setting a positive correlation between internal control quality and the information content of CSR reports by reducing companies’ agency costs. In line with that, Wang and Hu (2023) showed internal control effectiveness favorable for environmental information disclosure. Other authors came to similar conclusions (R. Huang & Huang, 2020; J. Zhang et al., 2024). With regard to environmental performance, ICS can have a positive effect on environmental performance through its positive effect on environmental investments (Liu et al., 2024). In line with these findings, the probability of ICS weaknesses is negatively related to ESG performance (Moffitt et al., 2023). Moreover, a company’s ESG rating is positive related to the growth (change of costs) of the ICS (Harasheh & Provasi, 2023), which emphasizes the relevance of the ICS for CSR performance. In contrast, there are hardly any studies that show negative effects of ICS. Only one study indicated no effect be-

tween ICS and CSR activities (Qin, 2019, cited by Li, 2020).

As for the ERM, a mediating or moderating effect was investigated in some studies. The ICS partially mediates the positive correlation between CSR and financial performance (L. Zhang & Su, 2023), or moderates the positive relationship of corporate environmental responsibility and financial performance (Liu et al., 2024). Other authors showed that the positive relationship between ESG ratings and financial performance was negatively moderated by ICS weaknesses (Boulhaga, Bouri, et al., 2023). Further studies showed additional effects of the ICS in relation to sustainability, such as that in environmental uncertainty the ICS effectiveness is higher (Jokipii, 2010). In light of the theoretical background of this paper, Boulhaga, Elbardan, and Elmassri (2023) found that a higher quality ICS (through the integration of CSR) can reduce conditional accounting conservatism, which helps to reduce agency conflicts (Shen et al., 2020). CSR and ICS can also have a positive impact on stakeholders view of the firm (Akisik & Gal, 2017).

With regard to German companies and the capital market, there is less research in the analysis at the intersection of ERM, ICS and sustainability. Based on a survey of 100 German industrial companies (including 65% listed companies), Scheffler and Flath (2023) found a full integration of sustainability risks into the ERM by 39%. Another study examined the statements on the appropriateness and effectiveness of the RMS and ICS for 2022, and observed only single statements on the still expandable integration of sustainability into the systems (Teucher & Ratzinger-Sakel, 2024). The Accounting Standards Committee of Germany (ASCG) examined the implementation of the ESRS with the help of a survey among the DAX40 in 2023. They found that 30% linked sustainability issues to risk management and 12.5% to the internal control system (DRSC, 2023). The results differ from two studies carried out by the consulting firm Stakeholder Reporting. These studies examined DAX40 sustainability reports for 2021 and 2022, with a partial aspect of the analysis focusing on the risk management system. The authors found that 85% in 2021 and 87.5% in 2022 of companies reported integrating sustainability risks into their risk management system (Stakeholder Reporting, 2022, 2023), although the description of the method and a theoretical embedding was barely provided.

The existing studies in relation to the German capital therefore fail to examine the company reports in relation to the ICS and RMS in more detail, or have methodological or scientific limitations. Against this background, this paper aims to answer the following main research question:

RQ1: What is the status of reporting on the sustainability-related ERM and ICS in German DAX companies?

Previous research emphasize that companies operating in environmental sensitive sectors are expected to disclose more environmental information (Chelli et al., 2018). As mentioned before, the second objective of this work is to examine

how these findings can be further developed in the area of reporting on sustainability-related ERM and ICS. This paper adopts the definitions for sensitive industries from previous studies (Garcia et al., 2017; Richardson & Welker, 2001). From the perspective of stakeholder agency theory, stakeholders may hold higher expectations of these organizations, for example, in the area of sustainable corporate governance or reporting. To meet the increased demand for information and to counteract the subsequent information imbalances, more detailed reporting may be necessary. Based on this, the second research question is as follows:

RQ2: How does the reporting differ between companies from environmental sensitive, and less sensitive industry sectors?

4. Research methodology

4.1. Sample selection and data sources

This study examines 29 German companies listed on the DAX40 for the period from 2022 and 2023. The corresponding DAX40 composition is based on the official report of the financial services provider Qontigo as of 31 December 2023.

The German capital market was chosen due to recent European (e.g. CSRD) and national regulatory (e.g. GCGC) developments. In addition to the substantive regulatory requirements, corporates operating in the German capital market are subject to two different forms of legal binding (hard law vs. soft law), making them a unique research object. DAX40 companies were selected as they represent the largest companies within the German capital market. In light of the theoretical framework, these organizations are expected to need and have a more developed RMS and ICS (Gatzert & Martin, 2015; Lundqvist, 2015), as they are facing a broader range of stakeholders (Hernández-Madrigal et al., 2020) which are increasing demanding sustainability disclosure (Boiral et al., 2020), and are more likely to face significant agency problems (Jensen & Meckling, 1976; Sarens & Christopher, 2010). In addition, they are subject to the most comprehensive regulatory provisions due to their size.

The observation period starts with the financial year 2022, as the revised version of GCGC had to be applied here for the first time. Moreover, the financial year 2023 is the last year before the application of the CSRD. Beyond that, the years 2022 and 2023 are the most recent years for which the data sources were available. The two-year period allows the observation of a possible trend, albeit only to a limited extent.

Data was sourced from annual reports, sustainability reports, if available integrated reports, which were manually downloaded from the respective companies' websites. These data sources were chosen because all DAX 40 companies publish such reports, and all data was expected here for the further course of the data analysis. Other published disclosure

vehicles such as "ESG presentations" and the like were not included. For reasons of comparability, only reports in English language were analyzed.

The initial sample consists of all 40 companies listed in the DAX40. This paper applies three exclusion criteria. First, eight financial and insurance companies are excluded due to their special regulations (Velte, 2017), following the approach of previous research (e.g. Deumes and Knechel, 2008; Gad, 2020). Second, for reasons of comparability and due to the research focus, this paper excludes two additional companies that do not report in accordance with German law. This approach is also in line with prior studies (Teucher & Ratzinger-Sakel, 2024). Third, one company was excluded due to its practical reporting practice. The non-financial statement according to GCC represents a relevant report element in the context of the data analysis (see Chapter 4.4). Therefore, one company, whose non-financial statement was integrated into another company's group statement, was excluded. The final sample therefore consists of 29 companies.

In light of the second research questions of this paper, the final sample is further categorized into companies operating in (non-) environmental sensitive industries.

Table 1 provides an overview of the selection process, distribution by industry as well as the proportion of companies operating in (non-) environmental sensitive sectors. Table B2 in the Appendix provides a comprehensive overview of all the companies examined, their associated industry and the company reports that were used for the data analysis.

4.2. Development of the disclosure index

A cornerstone of this research was the development of a suitable method for analyzing ESG reports in terms of the integration of sustainability into the ERM and ICS. This paper applies a self-developed disclosure index, which can be seen as a sub-type of a content analysis (Hassan & Marston, 2019; Krippendorff, 2019). In previous research, disclosure indices were often deployed to analyze disclosures on the ICS (e.g. Deumes and Knechel, 2008; Gad, 2020; Van De Poel and Vanstraelen, 2011). However, there is less use of indices with regard to the RMS. To the best of the author's knowledge, there has been no previous study that has examined the ERM and ICS in an integrated index. The attempt of this paper to develop an index for the integrated analysis of both systems therefore represents a novel approach.

The selection of the items is one of the key issues in the development of an index (Marston & Shrives, 1991). Against this background, the content of the index in this paper was developed on the basis of two popular conceptual frameworks by COSO. Specifically, it builds on the Enterprise Risk Management – Integrating with Strategy and Performance framework (see COSO, 2017), and the Internal Control – Integrated Framework (see COSO, 2013). These frameworks are the dominant models for both the RMS and ICS (Chen et al., 2017; Hayne & Free, 2014; Hernández-Madrigal et al., 2020). Previous research also used COSO frameworks for the development of disclosure indices (e.g. Hooghiemstra

Table 1: Sample selection and description.

Panel A: Sample size	
Companies in the DAX40	40
less	
Finance and insurance companies ^a	8
Not reporting to German law	2
Practical reporting practice	1
Final sample	29
Panel B: Distribution by industry sector	
Industry sector^b	Number of companies
Automobile	5
Chemicals	5
Consumer	3
Industrial	5
Pharma & Healthcare	3
Utilities	2
Others	6
Panel C: Categorization by (non-) environmental sensitive sector^c	
Companies in environmental sensitive sectors	13
Companies in non-environmental sensitive sectors	16

^a This includes two companies that were classified as financial holdings by the Frankfurt Stock Exchange and were therefore excluded.

^b The sectors were determined on the basis of the Frankfurt Stock Market classification. Sectors that only include one company (Construction, Retail, Software, Technology, Telecommunications, Transportation & Logistics) have been grouped as "Others".

^c Environmentally conscious sectors include the following industry sectors: Chemicals, Industrial, Utilities, Construction.

et al., 2015; Michelon et al., 2015), thereby supporting the approach of this paper.

The development of the index was based on the different conceptual elements of the frameworks. COSO defines so-called *components*, *principles*, and for the ICS framework *points of focus*. Principles represent the fundamental concepts within each component of the respective systems. Points of focus provide further explanations. Thus, the assessment criteria of the index are mainly derived from the principles, as they are well suited as evaluation items due to their substantial and practical focus. In some cases, the assessment criteria refer to the points of focus. Recently, COSO has further developed the internal control framework for sustainability reporting (see COSO, 2023). Similarly, the COSO risk management framework was adopted for the application to ESG-related risk (see COSO and wbcSD, 2018).

In doing so, COSO does not define any new components or principles, but interprets and explains them in the context of sustainability. With regard to the focus of this paper, the new interpretations are also considered when developing the assessment criteria. In addition to the COSO frameworks, the substantive structure of the index is also derived from the regulatory environment. This approach makes it possible to combine the more practical, process-related character of the COSO frameworks with some of the content-related focal points on sustainability issues found in recent legislation.

Notably, some of the components and associated principles in the frameworks are very similar (Lundqvist, 2015), for example in terms of their substantive focus and possible practical application. Therefore, certain principles are merged within the index to reflect these overlaps and to avoid double data collection. As mentioned before, this integrated approach for analyzing the reporting on the ERM and ICS represents a new method. Previous studies often examined either and exclusively the ERM or the ICS based on its components defined by COSO (e.g. Chen et al., 2017; Crawford and Nilsson, 2023; Michelon et al., 2015). Nonetheless, the merging of these frameworks is supported by recent literature, which emphasizes the substantive overlap and a sometimes unclear distinction between the individual principles (Prewett & Terry, 2018). The consolidation of the principles is supported not only by factual similarities but also by the integrated understanding of the ERM and ICS of this paper. In this way, the integrated index developed for this study will analyze the reporting of both systems in a connected and integrated way.

4.3. Integrated ERM and ICS disclosure index

The developed index is conceptually structured as follows. *Assessment criteria* form the specific items for the evaluation of each ESG report. These criteria are grouped into different *categories* based on their substantive focus. *Sections* then divide these categories from an organizational perspective.

The index is divided into three sections: (1) *General characteristics*, (2) *Performance on environmental issues*, and (3) *Monitoring*. The first section outlines general features of the systems from an organizational perspective. It includes the categories *System environment* (category 1), *Statement on system setup* (category 2) and *Specific system setup* (category 3). The second section provides an in-depth analysis of the practical application of the systems within the five topic-related ESRS for environmental aspects (ESRS E1 to E5). In this way it combines the process-related character of the COSO frameworks with content requirements of the regulatory environment. Consequently, the focus is less on the specific systems but more on the practical management of sustainability risks. Based on the five ESRS topics, this section consists of five categories *Performance I - V* (category 4 to 8). The third section deals with the monitoring of the systems and includes the categories *Internal oversight* (category 9) and *External oversight* (category 10).

Due to the different number of assessment criteria and index categories, the index is not weighted equally. Since in the later analysis each index section is considered separately and the different sections are not set in relation, a balancing weighting was not considered necessary.

All criteria are rated on a binary scale from 0 to 1. *Zero points* correspond to no statement on the respective criterion. *One point* correspond to the existence of the respective criterion. Here, for example, a detailed description of a process is necessary. Simply referring to the existence of a corresponding process is not considered sufficient. This leads to a total achievable score of 45 points (100%).

Table B3 in the Appendix provides a general overview of the index, including the allocation of the COSO framework principles and the embedded regulatory framework. Table B4 describes the derivation principles and a detailed description of the assessment criteria.

4.4. Data analysis

The data analysis comprises of the following four steps: selection of relevant disclosure vehicles, definition of report elements to be analyzed, performance of the assessment, and analysis of the data.

The *first step* was to select relevant disclosure instruments for the subsequent analysis. Drawing on the theoretical and regulatory framework, annual reports, sustainability reports, and integrated reports (where available) were considered relevant and defined as data sources. In a *second step*, the relevant report elements were determined. In light of the theoretical background, only those elements that are initiated by management were considered. Based on the regulatory framework, the entire management report and the sustainability report were considered relevant, thus serving as the fundamental database for the assessment. However, certain index categories deviate from this general logic. For index categories 2 and 9, only the corporate governance statement and the risk and opportunity report within the management report are considered. Both index categories were derived from the recommendations of the GCGC. The mentioned reporting elements represent the expected locations for corresponding reporting. For index category 10, the auditor's report is the relevant part of the report due to the focus of the assessment criteria. All other report components (e.g. the annual financial statement, or the remuneration report) were excluded from the analysis due to their irrelevance to the focus of this study. The general focus on the narrative elements of reports is also in line with the approach of previous studies (Hooghiemstra et al., 2015), as information about for example on the ICS (Deumes & Knechel, 2008; Van De Poel & Vanstraelen, 2011) and sustainability (Beattie, 2014) mainly occurs in these parts. The allocation of the analyzed report components to the assessment criteria can also be found in Table B3 in the Appendix. The *third step* was the assessment based on the criteria of the disclosure index. The final results were documented in a separate Excel spreadsheet. This spreadsheet forms the dependable data basis for the analysis. Simultaneously, the documents were coarsely systematically

coded according to the assessment criteria. MAXQDA2024 (Version 24.4.0) was utilized for the coding. The reason for the coding methodology was to achieve better transparency, traceability and thus finally the best possible consistency. In summary, a total of 2610 individual data points needed to be recorded for the entire sample and both years. The coding resulted in 2985 labeled passages. *Step four* was to analyze the collected data. In order to provide an overall statement about the reporting quality, this paper follows the four levels defined by Wulf et al. (2020). A score of 25% or less is considered *rudimentary*. Reporting is rated as *restrained* if the score is between 25% and 50%, and *satisfactory* if it is between 50% and 75%. Reports scoring over 75% are considered *comprehensive*. Percentages were calculated by dividing the individual points by the maximal achievable score. The presentation of the findings was partly inspired by the research of Braasch and Velte (2023) and Deumes and Knechel (2008), who also applied assessment models, although in a different research areas.

5. Results

5.1. Overall results

Table 2 shows the numerical scoring distribution with absolute and relative data for all 29 sample companies and both years analyzed.

In the following, the overall results are described. They are displayed in the *Total Score* column in Table 2. The percentages presented are calculated based on the maximum achievable score for the entire index. The analyzed companies achieved an average score (*Mean*) of 42% in 2022 and 45% in 2023. Both mean values are therefore in the range of restrained report quality. Although a positive trend can be identified, the *Mean Growth Rate* of +3.80% is only in the low single-digit percentage range. The maximum number of points (*Points_{Max}*) in both years is 67%, corresponding to a satisfactory level. On the other hand, the minimum number of points (*Points_{Min}*) is 22% in 2022 (20 % in 2023), and thereby in the area of rudimentary reporting quality. From the perspective of stakeholder agency theory, the results demonstrate that companies have significant potential for improvement in order to reduce information asymmetries and agency costs. The overall results are also presented in Figure 1.

5.2. Results by index section

This chapter describes the consolidated results per index section. The results for the three index sections are provided as *Sub-scores I-III* within Table 2. Base values for calculating the percentages are the respective maximum scores of the index sections. The results for the first index section (*General characteristics*) are above the mean of the total score in both years. The average sub-score increases from 69% in 2022 to 70% in 2023, equal to a mean growth rate of +1.24%. Both values are therefore in the satisfactory area of reporting quality. From a theoretical perspective, it can be implied

Table 2: Consolidated numerical scoring across sections and categories.

Year	(1) General characteristics				(2) Performance on environmental issues						(3) Monitoring			Total score	
	Cat. 1	Cat. 2	Cat. 3	Sub-score I	Cat. 4	Cat. 5	Cat. 6	Cat. 7	Cat.8	Sub-score II	Cat. 9	Cat. 10	Sub-score III		
2023	Score _A	4	4	4	12	6	6	6	6	30	5	3	8	50	
	Mean	3.79 (94)	2.90 (72)	1.66 (41)	8.34 (70)	3.34 (67)	1.24 (25)	2.66 (53)	1.83 (37)	1.97 (40)	11.03 (37)	0.79 (16)	0	0.79 (10)	20.17 (45)
	Median	4 (100)	3 (75)	2 (50)	8 (67)	3 (60)	1 (20)	3 (60)	2 (40)	2 (40)	11 (37)	0 (0)	0 (0)	0 (0)	20 (44)
	Std. Dev.	-	-	-	1.06	-	-	-	-	-	4.33	-	-	1.13	4.85
	Score _A	4	4	4	12	6	6	6	6	30	5	3	8	50	
2022	Mean (%)	3.97 (99)	2.69 (67)	1.59 (40)	8.24 (69)	3.14 (63)	1.24 (25)	2.38 (48)	1.55 (31)	1.62 (32)	9.93 (33)	0.55 (11)	0	0.55 (7)	18.72 (42)
	Median (abs)	4 (100)	3 (75)	2 (50)	8 (67)	3 (60)	1 (20)	2 (40)	1 (20)	9 (30)	0 (0)	0 (0)	0 (0)	18 (40)	
	Std. Dev.	-	-	-	1.22	-	-	-	-	-	4.16	-	-	0.89	4.66
	Mean Growth Rate	-	-	-	+1.24%	-	-	-	-	-	+5.51%	-	-	9.95%	+3.80%
	Sensitive sector	-	-	-	-	-	-	-	-	-	-	-	-	-	+3.11%
Non-sensitive sector	-	-	-	-	-	-	-	-	-	-	-	-	-	+4.39%	

Note: This table shows the total scoring, broken down by index section and the associated ten index categories. Moreover, it presents the results based on the sector categorization. Percentages have been rounded to one decimal place.
 Abbreviations: Score_A = Score achievable. Std. Dev. = Standard Deviation.

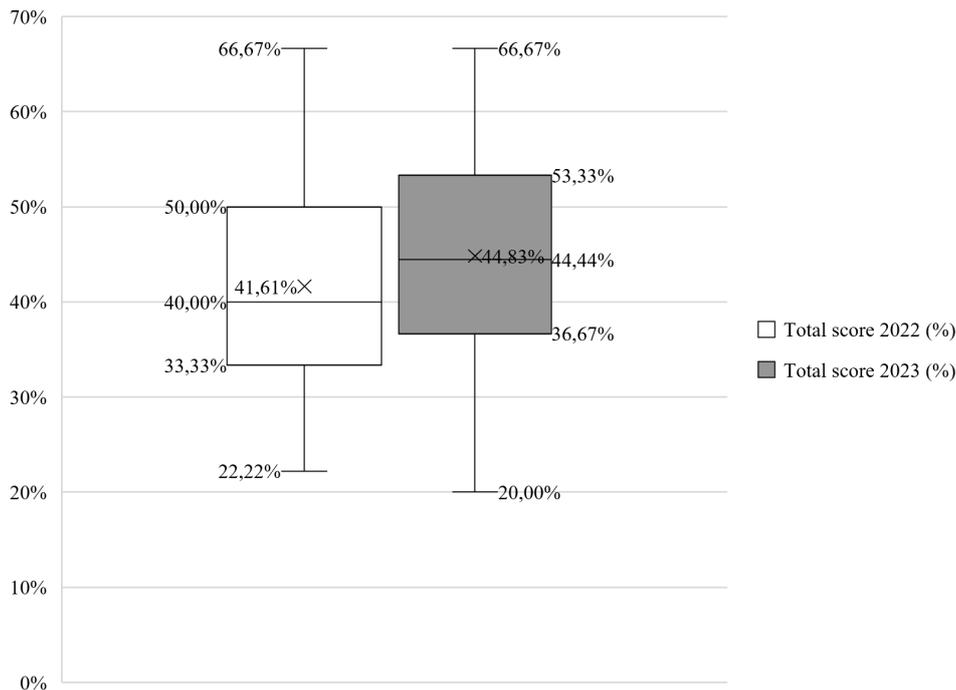


Figure 1: Overall scoring for 2022 and 2023.

that companies report in a comprehensive manner to meet stakeholder demands and reduce information asymmetries. The mean value for the second index section (*Performance on environmental issues*) is slightly below the respective value of the overall scoring. In 2022, the mean sub-score is 33% (37% in 2023). This places both years in the area of restrained reporting quality on performance on environmental issues. However, the mean growth rate (+5.51%) is slightly above the corresponding value of the overall scoring. The third index section (*Monitoring*) reveals the lowest sub-scoring. In 2022, the average is 7% and reaches 10% in 2023. This corresponds to a relatively high Mean Growth Rate of +9.95%. However, the values are clearly in the rudimentary range in terms of reporting quality. For the last two categories, there is the risk of insufficient reporting quality to adequately address information asymmetries between the management and external stakeholders.

5.3. Results by index category and criterion

The following chapter describes the results from the scoring categories and selected assessment criteria based on the mean values. For the index categories, Table 2 displays the consolidated results. Detailed results for all assessment criteria can be found in Table B5 in the Appendix.

Categories of the first index section. The first category (*General system environment*) addresses the system environment of the ERM and ICS, comprising relatively general items. This index category contains the highest mean value within the first index section for both years (2022: 99%; 2023: 95%). All four assessment criteria are close to the optimal mean of 1.0, indicating comprehensive report quality

and a solid approach to addressing information asymmetries. One possible explanation for these results may lie in the broad and generic focus of the specific COSO principles on which the assessment criteria in this category are defined. Category 2 (*Statement on system setup*) was developed on the basis of recommendation A.3 of the GCGC to integrate sustainability into the RMS and ICS and consequently into the CMS. For both years, the average mean values indicate satisfactory reporting quality and show a positive development (2023 = 72%; 2022 = 67%). When examining the declaration of compliance with recommendation A.3 of the GCGC (assessment criterion 2.1), a positive development can be observed, with all analyzed companies providing a corresponding statement in 2023. Thus, the maximum mean value of 1.0 was achieved in 2023 (2022: 0.90). One potential explanation for the lower score in 2022 might be that the GCGC revision was introduced during the financial year 2022. Therefore, not all companies may have published a corresponding declaration on the revised code in time. With regard to individual statements about the systems, three assessment criteria cover the RMS (assessment criterion 2.2), ICS (assessment criterion 2.3) and CMS (assessment criterion 2.4). The analyzed companies report most frequently on the integration of sustainability into the RMS (2023 = 0.90; 2022 = 0.86), followed by the ICS (2023 = 0.72; 2022 = 0.62), and lastly the CMS (2023 = 0.28; 2022 = 0.31). Except for the CMS, a positive development from 2022 to 2023 is noticeable. However, the values are distributed in the spectrum of restrained, satisfactory and comprehensive reporting quality, demonstrating uneven reporting quality on the systems. Particularly noteworthy is the significantly

lower level of reporting on the integration of sustainability into the CMS. The substantially lower score for the CMS may be explained by the fact that the GCGC only refers to the RMS and ICS in its recommendation A.3. However, the CMS is described as a part of the RMS and ICS in Principle 5 of the GCGC and should therefore also be included in the reporting. The results for the RMS are similar to those of Stakeholder Reporting (2023), who found that 87.5% of the analyzed DAX40 companies had integrated sustainability into their risk management system by 2022. Finally, category 3 (*Specific system setup*) deals with statements about selected system specifications. The analysis reveals restrained reporting quality for both years, with virtually no positive developments (2023 = 41%; 2022 = 40%). The analysis showed the highest mean values in the area of reporting on the application of environmental management systems (assessment criterion 3.4), which can be seen as part of a risk management systems (2023 = 0.93; 2022 = 0.93). Reporting on sustainability-related aspects of the respective industries (assessment criterion 3.1) received less attention and suggests restrained quality (2023 = 0.48; 2022 = 0.45). Companies barely report on their risk appetite (assessment criterion 3.2) or on the consideration of (un)intentional misuse (assessment criterion 3.3) in the area of sustainability issues. The mean values for both the reporting on the risk appetite (2023 = 0.10; 2022 = 0.07), as well as on the consideration of misuse (2023 = 0.14; 2022 = 0.14), clearly fall into the scope of rudimentary reporting.

Categories of the second index section. Categories 4 to 8 deal with practical statements on risk management performance for the five ESRS environmental issues and contain five assessment criteria each. Due to the large number of individual assessment criteria, the results are presented in condensed form in the following. A cross-category comparison of the mean values shows that the companies provide information particularly on the topics of climate change (2023 = 67%; 2022 = 63%), and water and marine resources (2023 = 53%; 2022 = 48%). Thus, reporting on both topics is of satisfactory reporting quality or has developed into this category during the observation period. The environmental topics of biodiversity and ecosystems (2023 = 37%; 2022 = 31%), and resource use and circular economy (2023 = 40%; 2022 = 32%) range in the middle of the five categories and demonstrate the biggest development. Nevertheless, companies only report on both topics with restrained quality on average. The topic of environmental pollution (2023 = 25%; 2022 = 25%) receives the least attention in the disclosure and is in the upper rudimentary range of reporting quality. Figure 2 shows the distribution of the scoring according to the five ESRS topics. As mentioned before, the same five assessment criteria are applied for each environmental ESRS topic (index category). Looking at the mean values for the five criteria across all ESRS topics, a clear focus emerges. Companies report most frequently on risk mitigation and response (2023 = 0.90; 2022 = 0.88), followed by risk identification, assessment and prioritization (2023 = 0.60; 2022 = 0.54). Reporting on the application of technology (2023

= 0.28; 2022 = 0.22), and the relevance and quality of data is almost identical (2023 = 0.24; 2022 = 0.22). It is particularly noteworthy that hardly any company makes a clear statement about integrating the ESRS topics into their risk management system (2023 = 0.18; 2022 = 0.12). This result is in line with the statements of Soomro and Lai (2017), who emphasize that the processes of sustainability and risk management are rarely integrated. However, the results for the integration of ESRS into the risk management system show a negative deviation from the results of a previous study. A survey found that 30% of DAX40 companies have integrated the ESRS into their RMS in 2023 (DRSC, 2023). This discrepancy highlights the necessity for companies to enhance their reporting quality in order to reduce potential information asymmetries. Figure 3 presents the average distribution of the assessment criteria across the five ESRS topics.

Categories of the third index section. Category 9 (*internal oversight*) focuses on internal oversight. Although this category achieves a better score than the second category in this section, it is still clearly in the area of rudimentary reporting quality (2023 = 16%; 2022 = 11%). The mean values of the assessment criteria are as follows. Companies most often report on the evaluation and improvement of systems in the context of sustainability (assessment criterion 9.3), albeit only in the area of rudimentary or restrained reporting quality (2023 = 0.38; 2022 = 0.28). The next most frequent item reported is the performance of an internal audit with a focus on ESG issues (assessment criterion 9.2). Although a relatively significant growth can be observed, both values indicate only rudimentary disclosure (2023 = 0.21; 2022 = 0.10). The remaining three items play a subordinate role in the reporting and score all in the area of restrained reporting quality. In light of existing research, the results of assessment criterion 9 are noteworthy. This item deals with the statement on the appropriateness and effectiveness of the ERM and ICS with regard to sustainability in accordance with recommendation A.5 of the GCGC. Teucher and Ratzinger-Sakel (2024) also analyzed the reports of the DAX 40 companies in 2022 for this statement. In their analysis of the inclusion of sustainability in the statement, they remain quantitatively imprecise, merely referring to "individual cases" (Teucher & Ratzinger-Sakel, 2024, p. 366). The analysis of this paper obtained a mean score for both years of 0.1. In absolute figures, this corresponds to three companies that have reported accordingly. However, due to the lack of precise information, it is not possible to say whether this research is completely in line with the previous findings. The items in category 10 (*external oversight*) analyze whether an external audit of the three systems ERM, ICS and CMS has been carried out with regard to sustainability on the basis of the IDW standards. No company has reported on a corresponding audit.

5.4. Results by categorization based on environmentally sensitive sectors

With regard to the second research question of this paper, the analysis revealed mixed results. Prior research suggests that companies operating in environmentally sensitive

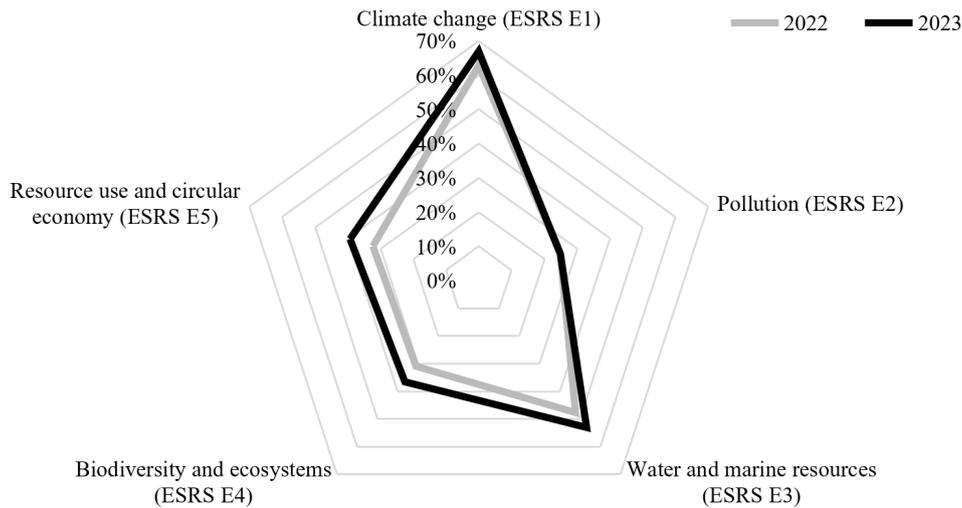


Figure 2: Scoring by index section two (ESRS topics).

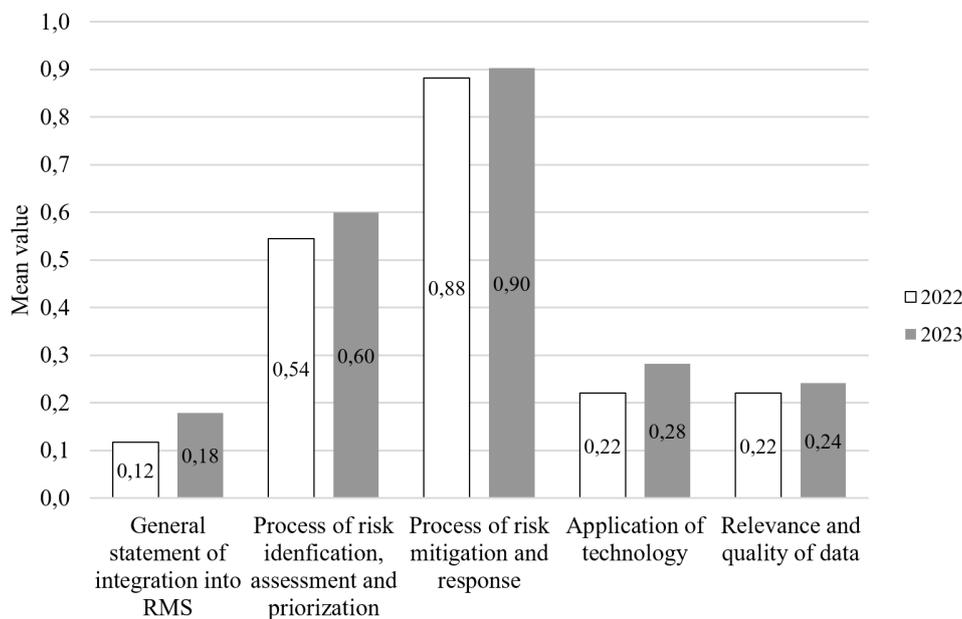


Figure 3: Scoring by assessment criterion in index section two.

sectors are expected to report more on sustainability (Chelli et al., 2018). From the perspective of stakeholder agency theory, external stakeholders may have higher expectations of these companies, for example with regard to sustainable corporate performance and reporting. Consequently, a more comprehensive reporting is required to address higher information asymmetries and reduce agency conflicts. For 2022, it can be observed that companies belonging to environmentally sensitive sectors produce better reporting quality. Nevertheless, the percentage difference is relatively marginal. Companies operating in the sensitive sector achieve a mean value of 44.72%, while the remaining companies achieve an average score of 40.28%. Both groups are still operating at

the level of restrained reporting quality. The trend towards 2023 shows that the quality of reporting by companies from non-sensitive sectors has improved by a greater proportion. The mean growth rate here is +4.39%, while the sensitive sector only gained +3.11% on average. Consequently, the statement that companies in environmental sectors produce better reporting quality can no longer be fully supported for 2023. The results imply that the affiliation to environmental sectors does not necessarily influence the reporting quality in the context of risk management and internal control systems. Figure 4 presents the findings according to the categorization for both years.

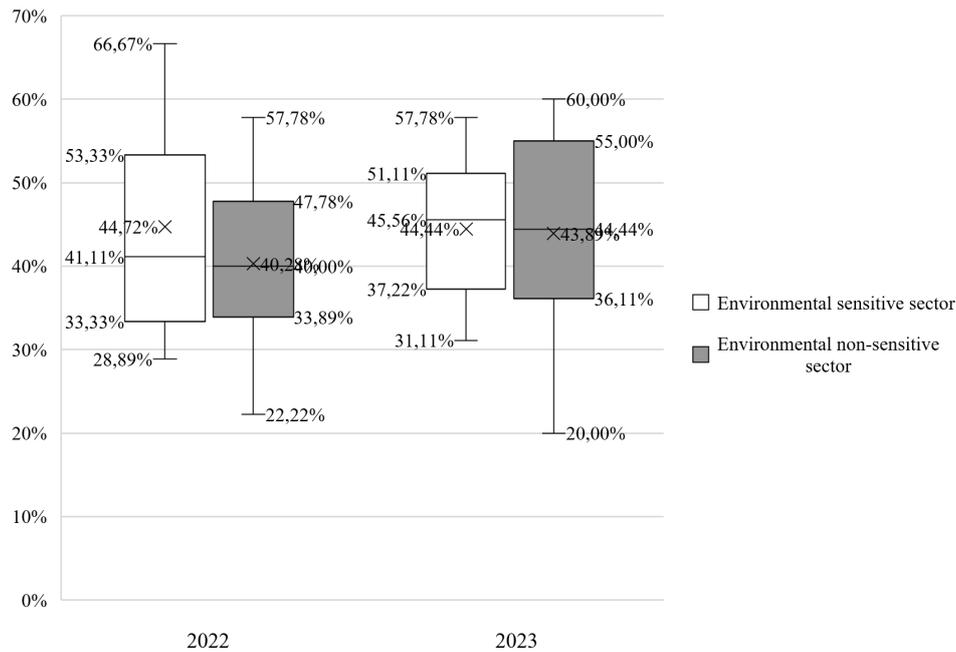


Figure 4: Scoring by (non-) environmental sensitive sector.

6. Conclusions

6.1. Implications for academia, policy and business practice

This study contributes to sustainable corporate governance literature and closes a research gap for the German capital market. The paper developed a disclosure index that examines both the risk management and internal control system in the context of sustainability. To the best of the author's knowledge, this integrated approach has never been applied in research before and thus represents a novelty in academic literature. When looking at the findings, the overall analysis showed that companies report with restrained quality on the integration of sustainability into their ERM and ICS. From the perspective of stakeholder agency theory, companies publish sustainability reports to address information asymmetries between the management and external stakeholders, helping to reduce agency costs (Buallay, 2022; Gray et al., 1995; Hahn & Kühnen, 2013; Michelin & Parbonetti, 2012). High-quality reporting on the systems can therefore be seen as an important approach for improving transparency about the way sustainability risks are managed or how the internal control system may be used to improve the reliability of sustainable reports. The analysis also revealed that some aspects of the systems in connection with sustainability are not reported on, or insufficiently reported on. Therefore, there is a probability of information asymmetries, which may lead to agency conflicts or higher agency costs. Due to scarce prior research, a discussion based on previous findings presented a hurdle. As presented in Chapter 5, some of the outcomes are in line with previous research. However, existing studies often lack substantive depth or methodological quality, thereby hindering a valid comparison. With regard to the second research objective

of this paper, the conclusions of earlier research cannot be clearly confirmed. Previous authors stated that companies operating in environmental sectors may publish more environmental information (e.g. Chelli et al., 2018). This does not necessarily appear to be the case when reporting on the interface between sustainability and ERM and ICS. While in 2022 a slightly better reporting quality from companies in environmentally sensitive sectors could still be observed, the results for 2023 showed an almost balanced reporting quality.

The results of this research also have implications for policy makers and regulatory bodies. The analysis showed a potential influence of specific legal provisions on reporting quality. For example, index category 3 was developed on the basis of the soft-law GCGC recommendation A.3. According to this, the ERM and ICS should also cover sustainability-related goals. Simultaneously, §161 GCC refers to the GCGC recommendations and requires a statement of compliance with the recommendations. If no compliance statement is provided, deviations must be explicitly explained. The results in category 3 reveal that for both years almost all companies (2023 = 100%, 2022 = 90%) did not declare any explicit deviation from recommendation A.3. The lower value for 2022 could possibly be explained by the implementation of the revised version within the observation period. However, further findings in category 3 reveal that fewer companies disclose a voluntary statement about the integration of sustainability into their systems. Under the stakeholder agency paradigm, simply stating compliance without providing specific substantial reference may not be as effective as making an explicit qualitative statement to reduce information asymmetries. This paper argues in favor of amendments to existing regulatory

obligations, also to counteract the risk of information asymmetries and agency conflicts. The goal should be a clear regulatory framework that enables companies to report in a reliable, comprehensive and consistent form on the integration of sustainability into the ERM and ICS. Consequently, distinct substantive requirements should be set. In the light of the findings of this paper, the following selected aspects should be included in future legal amendments. First, there should be a precise description of which sustainability aspects are included in the ERM and ICS. Second, the role of further corporate governance mechanisms (e.g. IAS) should be described in this context. Third, in view of the results of assessment criterion 9.5, a detailed qualitative statement on the effectiveness and appropriateness should be required. Fourth, reporting should also cover the performance and results of an external audit on the two sustainability-related systems. In addition to the substantive requirements, the legislation should also prescribe a clear location for such disclosures. From the author's point of view, the corporate governance statement may be suitable. A corresponding regulatory requirement may be integrated into Section 289 (4) GCC in conjunction with Section 289f HGB. According to the first paragraph, companies are already required to report on the key features of both systems, but in the context of the financial accounting process. The second paragraph requires the submission of the corporate governance statement.

Moreover, this paper reveals implications for business practice. The results indicate clear weaknesses (e.g. description monitoring aspects) when reporting on the integration of sustainability into the ERM and ICS. Consequently, companies may draw on these results, adapt their reporting and thus take a step towards reducing information asymmetries with external stakeholders.

6.2. Limitations and recommendations

This study holds some methodological limitations that also show the potential for future research. First, limitations exist with regard to the *sample selection*. This paper analyzed a narrow sample of 29 companies listed in the German DAX40. Thus, future research could also focus on the reporting practices of the financial and insurance companies that were excluded from this study. Furthermore, companies listed on other indices of the German capital market (e.g. MDAX, SDAX) could be examined. Since the obligations of the CSRD will also have to be applied by non-listed German companies in the future, an investigation may also be interesting here. Considering non-listed companies is also interesting since the recommendations of the GCGC to integrate sustainability into the RMS and ICS primarily target listed companies. Accordingly, these companies are required to submit a declaration of conformity. A comparison between listed and non-listed companies would also be interesting to examine the influence of recommendations and conformity declarations on disclosure practices. Moreover, the analysis could be carried out on the European capital market in order to examine possible differences between different national regulatory regimes on the reporting behavior and qual-

ity. Second, there are limitations in terms of the used *data sources*. The reports examined represent the dominant disclosure vehicles and are therefore suitable for the objective of this research using a disclosure index. However, it must be noted that not all written information necessarily reflects the actual integration of sustainability. Apart from text analysis, expert interviews or questionnaires for example could be additional qualitative methods to analyze the integration of sustainability into the RMS and ICS. Third, the application of a *disclosure index* as a type of content analysis holds some limitations. A disclosure index can only measure the content based on the pre-defined evaluation items. The development of the index used in this paper was given particular importance and has therefore also been developed based on the widely applied COSO frameworks for both systems. However, it cannot be ruled out that the index does not cover some aspects that were reported on and are also important for a further understanding of the integration of sustainability into the RMS and ICS. As mentioned by Marston and Shrivs (1991), the problem may arise that certain items are not as relevant for some companies as they are for others. Since this paper examines companies from different industry sectors, this might also be a weakness of the method. Furthermore, the index does not reflect the extent of reporting on a specific criterion. The sole use of a binary disclosure index makes it impossible to analyze where a particular assessment criterion was reported in the reports. Fourth, there are also limitations in terms of the *scope of this research*. Also, against the background of practical realizability, this paper focused on performance in relation to the environmental issues defined by the ESRS. Future research could build on the conceptual framework of this paper and investigate the reporting on the integration of social topics into the RMS and ICS, for example using the social issues defined by the ESRS. Fifth, there are limitations with regard to the *practical execution* of this study. Through the clear definition of the assessment criteria and the use of a coding tool, an attempt was made to maximize reproducibility. However, as this study was conducted by a single person, biases or individual imprecision cannot be ruled out.

7. Summary

The aim of this paper was an empirical-qualitative assessment of the reporting on the integration of sustainability into the RMS and ICS. The overarching research objective of this study, stated by the first research question, was to gain empirical insights into the quality of reporting. A second research question focused on potential disparities in the reporting quality between companies operating in (non-) environmental sensitive sectors.

Building on stakeholder agency theory, the study focused on a sample of 29 German DAX40 companies. For the analysis, a self-developed disclosure index was applied to ESG reports for the financial years 2022 and 2023. The approach of integrating both systems into one index for an analysis of

ESG reporting represents a new method in academic sustainable corporate governance literature.

The topic, sample and observation period was motivated by societal, practical and regulatory relevance. Organizations are facing a growing number of complex sustainability issues and the associated risks, while stakeholders are increasingly demanding sustainable corporate performance and associated reporting. Integrating sustainability into the ERM and ICS and corresponding reporting can therefore be an approach to adequately address sustainability risks, improve sustainable corporate performance and increase the reliability of sustainability reports. Furthermore, companies are subject to an increasing number of current and future regulations at both the national and European level.

This research contributes to an identified research gap. There is very scarce qualitative research at the intersection between sustainability and risk management and internal control systems. Existing research mainly uses quantitative methods. With regard to the German capital market, previous qualitative research pays little attention to sustainability or lacks methodical quality. To the best of the author's knowledge, there is no research examining the integration of sustainability into ERM and ICS with a combined disclosure index. Through the lens of stakeholder agency theory, reporting on these systems can help to reduce information asymmetries between the management and external stakeholders.

The empirical results indicate an overall restrained reporting quality. There are significant differences among the individual index sections, categories and assessment criteria. A minor trend was observed from 2022 to 2023. Therefore, with regard to the first research question, the results show a mixed picture. In light of the theoretical framework, there is a risk of information asymmetries between management and external stakeholders, which can lead to increased agency costs. As for the second research question, in 2022 it can be noted that companies in environmentally sensitive sectors are reporting at a higher rate. In 2023, this observation can no longer be supported.

The research has several methodological limitations, such as the size of the sample, possible methodological deficiencies in the development of the index, or personal bias when conducting the analysis. Based on a number of limitations, recommendations for future research can be derived.

The research is highly relevant for business practice, academia and regulatory bodies. The results show distinct shortcomings that can be drawn upon by policy makers and companies in the future.

References

- Akisik, O., & Gal, G. (2017). The impact of corporate social responsibility and internal controls on stakeholders' view of the firm and financial performance. *Sustainability Accounting, Management and Policy Journal*, 8, 246–280.
- Anderson, D., Francis, J. R., & Stokes, D. J. (1993). Auditing, directorships and the demand for monitoring. *Journal of Accounting and Public Policy*, 12, 353–375.
- Anderson, D. R., & Anderson, K. E. (2009). Sustainability Risk Management. *Risk Management and Insurance Review*, 12, 25–38.
- Bartuschka, W. (2022). Angemessenheit und Wirksamkeit von Systemen der internen Unternehmensüberwachung im Kontext von FISG und DCGK 2022. *Betriebs-Berater*, 1387–1390.
- Beattie, V. (2014). Accounting narratives and the narrative turn in accounting research: Issues, theory, methodology, methods and a research framework. *The British Accounting Review*, 46, 111–134.
- Berwanger, J., & Kullmann, S. (2012). *Interne Revision: Funktion, Rechtsgrundlagen und Compliance*. Springer Fachmedien Wiesbaden.
- Boiral, O., Guillaume, L., Heras-Saizarbitoria, I., & Tayo Tene, C. V. (2018). Adoption and Outcomes of ISO 14001: A Systematic Review. *International Journal of Management Reviews*, 20, 411–432.
- Boiral, O., Talbot, D., & Brotherton, M.-C. (2020). Measuring sustainability risks: A rational myth? *Business Strategy and the Environment*, 29, 2557–2571.
- Boulhaga, M., Bouri, A., Elamer, A. A., & Ibrahim, B. A. (2023). Environmental, social and governance ratings and firm performance: The moderating role of internal control quality. *Corporate Social Responsibility and Environmental Management*, 30, 134–145.
- Boulhaga, M., Elbardan, H., & Elmassri, M. (2023). The effect of internal control and corporate social responsibility on conditional accounting conservatism: Evidence from France. *Journal of Corporate Accounting & Finance*, 34, 228–241.
- Braasch, A., & Velte, P. (2023). Climate reporting quality following the recommendations of the task force on climate-related financial disclosures: A Focus on the German capital market. *Sustainable Development*, 31, 926–940.
- Brennan, N. (2005). Corporate governance: Accountability, enterprise and international comparisons, 2005. *The International Journal of Accounting*, 40, 425–428.
- Brown, I., Steen, A., & Foreman, J. (2009). Risk Management in Corporate Governance: A Review and Proposal. *Corporate Governance: An International Review*, 17, 546–558.
- Buallay, A. M. (2022). Relevant Theories to Sustainability Reporting. In *International Perspectives on Sustainability Reporting* (pp. 61–76). Emerald Publishing Limited.
- Capelli, P., Ielasi, F., & Russo, A. (2021). Forecasting volatility by integrating financial risk with environmental, social, and governance risk. *Corporate Social Responsibility and Environmental Management*, 28, 1483–1495.
- Chairani, C., & Siregar, S. V. (2021). The effect of enterprise risk management on financial performance and firm value: the role of environmental, social and governance performance. *MEDAR*, 29, 647–670.
- Chan, K. C., Chen, Y., & Liu, B. (2021). The Linear and Non-Linear Effects of Internal Control and Its Five Components on Corporate Innovation: Evidence from Chinese Firms Using the COSO Framework. *European Accounting Review*, 30, 733–765.
- Chelli, M., Durocher, S., & Fortin, A. (2018). Normativity in Environmental Reporting: A Comparison of Three Regimes. *Journal of Business Ethics*, 149, 285–311.
- Chen, H., Dong, W., Han, H., & Zhou, N. (2017). A comprehensive and quantitative internal control index: construction, validation, and impact. *Review of Quantitative Finance and Accounting*, 49, 337–377.
- Collier, P. M. (2008). Stakeholder accountability: A field study of the implementation of a governance improvement plan. *Accounting, Auditing & Accountability Journal*, 21, 933–954.
- COSO. (2013). Internal Control — Integrated Framework Executive Summary. Retrieved April 4, 2024, from https://www.coso.org/_file/ugd/3059fc_1df7d5dd38074006bce8df621a942cf.pdf
- COSO. (2017). Enterprise Risk Management Integrating with Strategy and Performance. Retrieved February 4, 2024, from https://aaahq.org/portals/0/documents/coso/coso_erm_2017_main_v1_20230815.pdf
- COSO. (2023). Achieving Effective Internal Control over Sustainability Reporting (ICSR): Building Trust and Confidence through the COSO Internal Control—Integrated Framework. Retrieved February 4, 2024, from <https://theia.se/wp-content/uploads/2023/04/COSO-ICSR-Report.pdf>
- COSO and wbcscd. (2018). Enterprise Risk Management Applying enterprise risk management to environmental, social and governance-

- related risks. Retrieved January 4, 2024, from https://docs.wbcsd.org/2018/10/COSO_WBCSD_ESGERM_Guidance.pdf
- Crawford, J., & Jabbour, M. (2024). The relationship between enterprise risk management and managerial judgement in decision-making: A systematic literature review. *International Journal of Management Reviews*, 26, 110–136.
- Crawford, J., & Nilsson, F. (2023). Integrating ESG Risks into Control and Reporting: Evidence from Practice in Sweden. In T. Rana, J. Svanberg, P. Öhman, & A. Lowe (Eds.), *Handbook of Big Data and Analytics in Accounting and Auditing* (pp. 255–277). Springer Nature.
- Deumes, R., & Knechel, W. R. (2008). Economic Incentives for Voluntary Reporting on Internal Risk Management and Control Systems. *AUDITING: A Journal of Practice & Theory*, 27, 35–66.
- Dienes, D., & Velte, P. (2016). The Impact of Supervisory Board Composition on CSR Reporting. Evidence from the German Two-Tier System. *Sustainability*, 8, 63.
- Dobler, M., Lajili, K., & Zéghal, D. (2014). Environmental Performance, Environmental Risk and Risk Management. *Business Strategy and the Environment*, 23, 1–17.
- DRSC. (2023). European Sustainability Reporting Standards (ESRS): DRSC's Survey on the Implementation of ESRS at the German DAX 40 Companies. Retrieved March 21, 2024, from https://www.drsc.de/app/uploads/2023/09/20230929_Short-Report_DAX-40-Companies_ESRS-Implementation.pdf
- Dumay, J., & Hossain, M. A. (2019). Sustainability Risk Disclosure Practices of Listed Companies in Australia. *Australian Accounting Review*, 29, 343–359.
- Earle, T. C. (2009). Trust, Confidence, and the 2008 Global Financial Crisis. *Risk Analysis*, 29, 785–792.
- European Commission. (2021). The European Green Deal - Striving to be the first climate-neutral continent - European Commission. Retrieved May 10, 2024, from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en
- Fakir, A. N. M. A., & Jusoh, R. (2020). Board Gender Diversity and Corporate Sustainability Performance: Mediating Role of Enterprise Risk Management. *The Journal of Asian Finance, Economics and Business*, 7, 351–363.
- Fernandez-Feijoo, B., Romero, S., & Ruiz, S. (2014). Effect of Stakeholders' Pressure on Transparency of Sustainability Reports within the GRI Framework. *Journal of Business Ethics*, 122, 53–63.
- Foerstl, K., Reuter, C., Hartmann, E., & Blome, C. (2010). Managing supplier sustainability risks in a dynamically changing environment-Sustainable supplier management in the chemical industry. *Journal of Purchasing and Supply Management*, 16, 118–130.
- Freeman, R. E. (1984). *Strategic management: a stakeholder approach*, Pitman series in business and public policy. Pitman.
- Freeman, R. E. (2010). *Strategic Management: A Stakeholder Approach*. Cambridge University Press.
- Gad, J. (2020). Voluntary disclosures on control system over financial reporting and corporate governance mechanisms: Evidence from Poland. *Journal of East European Management Studies*, 25, 698–729.
- Garcia, A. S., Mendes-Da-Silva, W., & Orsato, R. J. (2017). Sensitive industries produce better ESG performance: Evidence from emerging markets. *Journal of Cleaner Production*, 150, 135–147.
- Gatzert, N., & Martin, M. (2015). Determinants and Value of Enterprise Risk Management: Empirical Evidence From the Literature. *Risk Management and Insurance Review*, 18, 29–53.
- Giannakis, M., & Papadopoulos, T. (2016). Supply chain sustainability: A risk management approach. *International Journal of Production Economics*, 171, 455–470.
- Government Commission on the German Corporate Governance Code. (2022). German Corporate Governance Code. Retrieved April 20, 2024, from <https://www.dcgk.de/en/code/current-version/a-management-and-supervision.html>
- Gray, R., Kouhy, R., & Lavers, S. (1995). Corporate social and environmental reporting: a review of the literature and a longitudinal study of UK disclosure. *Accounting, Auditing & Accountability Journal*, 8, 47–77.
- Hahn, R., & Kühnen, M. (2013). Determinants of sustainability reporting: a review of results, trends, theory, and opportunities in an expanding field of research. *Journal of Cleaner Production*, 59, 5–21.
- Harasheh, M., & Provasi, R. (2023). A need for assurance: Do internal control systems integrate environmental, social, and governance factors? *Corporate Social Responsibility and Environmental Management*, 30, 384–401.
- Hassan, O. A. G., & Marston, C. (2019). Corporate Financial Disclosure Measurement in the Empirical Accounting Literature: A Review Article. *The International Journal of Accounting*, 54, 1950006.
- Hayne, C., & Free, C. (2014). Hybridized professional groups and institutional work: COSO and the rise of enterprise risk management. *Accounting, Organizations and Society*, 39, 309–330.
- Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31, 405–440.
- Henk, O. (2020). Internal control through the lens of institutional work: a systematic literature review. *Journal of Management Control*, 31, 239–273.
- Hernández-Madrugal, M., Aibar-Guzmán, C., Aibar-Guzmán, B., & Ramírez-Flores, É. (2020). Are external pressures always behind ERM implementation? Evidence from Spanish listed firms. *International Journal of Disclosure and Governance*, 17, 86–100.
- Hill, C. W. L., & Jones, T. M. (1992). Stakeholder-Agency Theory. *Journal of Management Studies*, 29, 131–154.
- Hooghiemstra, R., Hermes, N., & Emanuels, J. (2015). National Culture and Internal Control Disclosures: A Cross-country Analysis. *Corporate Governance*, 23, 357–377.
- Huang, P., Jiao, Y., & Li, S. (2022). Impact of internal control quality on the information content of social responsibility reports: A study based on text similarity—Evidence from China. *International Journal of Accounting Information Systems*, 45, 100558.
- Huang, R., & Huang, Y. (2020). Does Internal Control Contribute to a Firm's Green Information Disclosure? Evidence from China. *Sustainability*, 12, 3197.
- IDW. (2017a). IDW Prüfungsstandard: Grundsätze ordnungsmäßiger Prüfung des internen Kontrollsystems des internen und externen Berichtswesens (IDW PS 982).
- IDW. (2017b). IDW Prüfungsstandard: Grundsätze ordnungsmäßiger Prüfung von Internen Revisionssystemen (IDW PS 983).
- IDW. (2017c). IDW Prüfungsstandard: Grundsätze ordnungsmäßiger Prüfung von Risikomanagementsystemen (IDW PS 981).
- IDW. (2022). IDW Prüfungsstandard: Grundsätze ordnungsmäßiger Prüfung von Compliance Management Systemen (IDW PS 980 n.F. (09.2022)).
- IDW. (2023). IDW Positionspaper: Compliance-Kultur in deutschen Unternehmen verbessern - Zur Empfehlung A.5 DCGK 2022.
- Jankensgård, H. (2019). A theory of enterprise risk management. *Corporate Governance: The International Journal of Business in Society*, 19, 565–579.
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3, 305–360.
- Jokipii, A. (2010). Determinants and consequences of internal control in firms: a contingency theory based analysis. *Journal of Management & Governance*, 14, 115–144.
- Jones, M. J. (2008). Internal control, accountability and corporate governance: Medieval and modern Britain compared. *Accounting, Auditing & Accountability Journal*, 21, 1052–1075.
- Kim, Y. S., Kim, Y., & Kim, H. (2017). Corporate Social Responsibility and Internal Control Effectiveness. *Asia-Pacific Journal of Financial Studies*, 46, 341–372.
- Krippendorff, K. (2019). *Content Analysis: An Introduction to Its Methodology*. SAGE Publications, Inc.
- Kuo, Y.-F., Lin, Y.-M., & Chien, H.-F. (2021). Corporate social responsibility, enterprise risk management, and real earnings management: Evidence from managerial confidence. *Finance Research Letters*, 41, 101805.

- Lenssen, J.-J., Dentchev, N. A., & Roger, L. (2014). Sustainability, Risk management and governance: towards an integrative approach. *Corporate Governance (Bingley)*, 14, 670–684.
- Li, X. (2020). The effectiveness of internal control and innovation performance: An intermediary effect based on corporate social responsibility. *PLOS ONE*, 15, 1–31.
- Liu, X., Liu, S., Wang, J., & Chen, H. (2024). Does internal control affect corporate environmental responsibility? Evidence from China [Ahead-of-print]. *International Journal of Emerging Markets*.
- Louie, J., Ahmed, K., & Ji, X.-D. (2019). Voluntary disclosures practices of family firms in Australia. *Accounting Research Journal*, 32, 273–294.
- Lu, H., Liu, X., & Falkenberg, L. (2022). Investigating the Impact of Corporate Social Responsibility (CSR) on Risk Management Practices. *Business & Society*, 61, 496–534.
- Lundqvist, S. A. (2015). Why firms implement risk governance – Stepping beyond traditional risk management to enterprise risk management. *Journal of Accounting and Public Policy*, 34, 441–466.
- Marston, C. L., & Shrivs, P. J. (1991). The use of disclosure indices in accounting research: A review article. *The British Accounting Review*, 23, 195–210.
- Mercer, M. (2004). How Do Investors Assess the Credibility of Management Disclosures? *Accounting Horizons*, 18, 185–196.
- Michelon, G., Bozzolan, S., & Beretta, S. (2015). Board monitoring and internal control system disclosure in different regulatory environments. *Journal of Applied Accounting Research*, 16, 138–164.
- Michelon, G., & Parbonetti, A. (2012). The effect of corporate governance on sustainability disclosure. *Journal of Management & Governance*, 16, 477–509.
- Moffitt, J. S., Patin, J.-C. A., & Watson, L. (2023). Corporate Environmental, Social, and Governance (ESG) Performance and the Internal Control Environment. *Accounting Horizons*, 1–22.
- Mohammed, M. (2000). The ISO 14001 EMS Implementation Process and Its Implications: A Case Study of Central Japan. *Environmental Management*, 25, 177–188.
- Musallam, S. R. M. (2018). The direct and indirect effect of the existence of risk management on the relationship between audit committee and corporate social responsibility disclosure. *Benchmarking: An International Journal*, 25, 4125–4138.
- Naciti, V., Cesaroni, F., & Pulejo, L. (2021). Corporate governance and sustainability: a review of the existing literature. *Journal of Management and Governance*, 26, 1–20.
- Naseem, T., Shahzad, F., Asim, G. A., Rehman, I. U., & Nawaz, F. (2020). Corporate social responsibility engagement and firm performance in Asia Pacific: The role of enterprise risk management. *Corporate Social Responsibility and Environmental Management*, 27, 501–513.
- Nawrocka, D., & Parker, T. (2009). Finding the connection: environmental management systems and environmental performance. *Journal of Cleaner Production*, 17, 601–607.
- Palousis, N., Luong, L., & Abhary, K. (2010). Sustainability risk identification in product development. *International Journal of Sustainable Engineering*, 3, 70–80.
- Pérez-Cornejo, C., & de Quevedo-Puente, E. (2023). How corporate social responsibility mediates the relationship between corporate reputation and enterprise risk management: evidence from Spain. *Eurasian Business Review*, 13, 363–383.
- Prewett, K., & Terry, A. (2018). COSO's Updated Enterprise Risk Management Framework—A Quest For Depth And Clarity. *Journal of Corporate Accounting & Finance*, 29, 16–23.
- Qin, D. S. (2019). Corporate Social Responsibility, Internal Control and Brand Value. *Value Engineering*, 38, 296–299.
- Rezaee, Z. (2016). Business sustainability research: A theoretical and integrated perspective. *Journal of Accounting Literature*, 36, 48–64.
- Richardson, A. J., & Welker, M. (2001). Social disclosure, financial disclosure and the cost of equity capital. *Accounting, Organizations and Society*, 26, 597–616.
- Sarens, G., & Christopher, J. (2010). The association between corporate governance guidelines and risk management and internal control practices: Evidence from a comparative study. *Managerial Auditing Journal*, 25, 288–308.
- Sarens, G., De Beelde, I., & Everaert, P. (2009). Internal audit: A comfort provider to the audit committee. *The British Accounting Review*, 41, 90–106.
- Sarkis, J. (2006). The adoption of environmental and risk management practices: Relationships to environmental performance. *Ann Oper Res*, 145, 367–381.
- Scheffler, R., & Flath, T. (2023). Risikomanagementsysteme Status in deutschen Industrieunternehmen in bewegten Zeiten. *Zeitschrift für Risikomanagement (ZfRM)*, 3.
- Shad, M. K., Lai, F.-W., Fatt, C. L., Klemeš, J. J., & Bokhari, A. (2019). Integrating sustainability reporting into enterprise risk management and its relationship with business performance: A conceptual framework. *Journal of Cleaner Production*, 208, 415–425.
- Shah, S. Q. A., Lai, F.-W., Shad, M. K., Hamad, S., & Ellili, N. O. D. (2024). Exploring the effect of enterprise risk management for ESG risks towards green growth [Ahead-of-print]. *International Journal of Productivity and Performance Management*.
- Sharfman, M. P., & Fernando, C. S. (2008). Environmental risk management and the cost of capital. *Strategic Management Journal*, 29, 569–592.
- Shen, X., Ho, K.-C., Yang, L., & Wang, L. F.-S. (2020). Corporate social responsibility, market reaction and accounting conservatism. *Kybernetes*, 50, 1837–1872.
- Short, H., Keasey, K., Wright, M., & Hull, A. (1999). Corporate governance: from accountability to enterprise. *Accounting and Business Research*, 29, 337–352.
- Soomro, M. A., & Lai, F.-W. (2017). Examining A New Paradigm of Enterprise Sustainability Risk Management. *Global Business and Management Research: An International Journal*, 9, 328–337.
- Stakeholder Reporting. (2022). CSRD-Readiness: Wie gut sind DAX40 Unternehmen aufgestellt? [WWW Document]. Retrieved June 19, 2024, from <https://www.stakeholder-reporting.com/csr-readiness-wie-gut-sind-dax40-unternehmen-aufgestellt/>
- Stakeholder Reporting. (2023). CSRD-Readiness im DAX40: Berichtsgrundlagen (Teil 1) [WWW Document]. Retrieved June 19, 2024, from <https://www.stakeholder-reporting.com/csr-readiness-im-dax40/>
- Stiglbauer, M., & Velte, P. (2014). Impact of soft law regulation by corporate governance codes on firm valuation: the case of Germany. *Corporate Governance*, 14, 395–406.
- Subramaniam, N., Wahyuni, D., Cooper, B. J., Leung, P., & Wines, G. (2015). Integration of carbon risks and opportunities in enterprise risk management systems: evidence from Australian firms. *Journal of Cleaner Production*, 96, 407–417.
- Teucher, C., & Ratzinger-Sakel, N. V. S. (2024). Angemessenheits- und Wirksamkeitsaussage zum internen Kontrollsystem und Risikomanagementsystem. *WPg*, 361–368.
- Traxler, A. A., Schrack, D., & Greiling, D. (2020). Sustainability reporting and management control – A systematic exploratory literature review. *Journal of Cleaner Production*, 276, 122725.
- Valinejad, F., & Rahmani, D. (2018). Sustainability risk management in the supply chain of telecommunication companies: A case study. *Journal of Cleaner Production*, 203, 53–67.
- Van De Poel, K., & Vanstraelen, A. (2011). Management Reporting on Internal Control and Accruals Quality: Insights from a "Comply-or-Explain" Internal Control Regime. *Auditing: A Journal of Practice & Theory*, 30, 181–209.
- Velte, P. (2017). Does ESG performance have an impact on financial performance? Evidence from Germany. *Journal of Global Responsibility*, 8, 169–178.
- Velte, P. (2022). Überwachung des internen Corporate Governance-Systems nach § 107 Abs. 3 Satz 2 AktG durch den Prüfungsausschuss: Zur normativen Stärkung der Zusammenarbeit zwischen Prüfungsausschuss, interner Revision und Abschlussprüfer bei börsennotierten Aktiengesellschaften. *Der Konzern: Zeitschrift für Gesellschaftsrecht, Steuerrecht, Bilanzrecht und Rechnungslegung der verbundenen Unternehmen*, 20, 275–289.
- Velte, P. (2023). Sustainable institutional investors and corporate biodiversity disclosure: Does sustainable board governance matter? *Corporate Social Responsibility and Environmental Management*, 30, 3063–3074.

- Wang, Y., & Hu, Y. (2023). An empirical study on the relationship between the internal control effectiveness and the quality of environmental information disclosure: public data from listed companies in China's oil and gas industry. *Journal of Computing and Electronic Information Management*, 11, 47–52.
- Wu, S. J., Melnyk, S. A., & Calantone, R. J. (2008). Assessing the Core Resources in the Environmental Management System From the Resource Perspective and the Contingency Perspective. *IEEE Transactions on Engineering Management*, 55, 304–315.
- Wu, T., & Blackhurst, J. (Eds.). (2009). *Managing Supply Chain Risk and Vulnerability: Tools and Methods for Supply Chain Decision Makers*. Springer London.
- Wulf, I., Friedrich, T. J., Senger, A., & Staikowski, R. A. (2020). Klimabezogene Angaben in der nichtfinanziellen Pflichtberichterstattung—Deskriptive Analyse und empirische Evidenz zur Berichtsqualität der DAX30-Unternehmen. *Zeitschrift für Umweltpolitik & Umweltrecht*, 4, 460–495.
- Zhang, J., Zhang, L., & Zhang, M. (2024). Media pressure, internal control, and corporate environmental information disclosure. *Finance Research Letters*, 63, 105369.
- Zhang, L., & Su, W. (2023). Corporate social responsibility, internal control, and firm financial performance. *Frontiers in Psychology*, 13, 977996.



A New Dimension of Transparency: ESG Disclosure and Its Effect on Shareholder Behavior

Tobias Keserü

Technical University of Munich

Abstract

This study examines the impact of Environmental, Social, and Governance Disclosure (ESGD) on a company's ownership structure in predominantly developed economies. It aims to assess whether ESGD influences the shares held by different investor types, with a focus on institutional investors. Using data from 2016 to 2022, ESGD is measured relatively and absolutely, while ownership is categorized into corporate, government, individual, and institutional types. A multivariate regression assesses the overall impact, and a univariate regression specifically examines the effect on institutional ownership. The analysis reveals a significant link between ESGD and ownership structure, suggesting that ESGD shapes ownership dynamics. In particular, institutional investors respond positively to relative ESGD, valuing how a company's ESG transparency compares to its peers. The study acknowledges limitations like the short time frame and potential biases in the database. Nevertheless, the findings suggest that companies can attract institutional investors by improving ESG transparency, even if actual ESG outcomes are modest. This research contributes to the understanding of how ESG transparency shapes investor behavior, offering valuable insights for companies, investors, and policymakers.

Keywords: corporate governance; ESG disclosure; institutional investors; ownership structure; transparency

1. Introduction

For decades, investor behavior has been portrayed as primarily driven by financial performance, with iconic figures such as Warren Buffett representing the prototypical investor focused on fundamental value metrics: key numbers about cash flows, returns, and market positioning have been central (Koller, 2020). The traditional notion of an investor conjures images of individuals or institutions making decisions based solely on the bottom line, often sidelining concerns related to non-financial factors such as environmental, social, and governance (ESG) practices (Zairis et al., 2024, p.

7). However, investment priorities shifted significantly in recent years: A growing number of investors increasingly value not only financial performance but also sustainable and responsible business practices, with ESG becoming more relevant in business and society (Zairis et al., 2024, p. 7). From individual daily habits to national and global strategic initiatives, sustainable and socially accepted behavior has emerged as a critical focus. In response to these broader societal shifts, corporates increasingly address ESG and its implications (Barnea & Rubin, 2010, p. 71). Consumers, employees, and regulators are pushing for greater corporate accountability on topics like climate change, social justice, and ethical governance. As a result, companies are under increasing pressure to demonstrate their commitment to ESG principles (Cao et al., 2019, p. 289). This, in turn, is directly connected to how firms communicate their ESG efforts to stakeholders, notably through **ESG disclosures** (ESGD). The rise of ESGD represents a new dimension of transparency, providing insights into a company's social and environmen-

I would like to express my sincere gratitude to my supervisor, Maximilian Nadicksbernd, for his continuous guidance, expert advice, and support throughout the entire thesis process. His insightful feedback greatly contributed to the development of this work. I also extend my heartfelt thanks to my family and friends for their constant encouragement and support. Their presence throughout this journey provided essential balance, helping me navigate challenges and recognize milestones with perspective and resilience.

tal performance in addition to its financial health. While ESGD is still a relatively new trend, its influence on investor behavior is becoming increasingly recognizable (Cao et al., 2019, p. 287). As key recipients of ESG information, investors use these insights to assess corporate responsibility and long-term sustainability, shaping their investment decisions accordingly (Grewal et al., 2019, p. 3061). Regulatory frameworks for ESGD are still evolving, leaving companies with considerable flexibility in determining the extent and depth of their disclosures. As Cao et al. (2019, p. 288) note, "among the firms that do commit to CSR reporting, there exists enormous variance in the level of disclosures." Some companies only briefly mention ESG initiatives in annual reports, while others produce comprehensive, standalone sustainability reports (Dhaliwal et al., 2011, p. 60). This variability, coupled with the largely voluntary nature of ESGD, raises questions about the utility of this form of transparency and its impact on **investor behavior**. One concern surrounding ESGD is the potential for information overload: with the vast amounts of data companies provide in their reports and the lack of shared norms, it can be challenging for investors to identify what is material and relevant to their decision-making process (D. M. Christensen et al., 2022, p. 164). This issue becomes even more complex as different investor types may interpret ESG information differently. Professional analysts, for instance, often have the resources and expertise to evaluate ESGD critically, while retail investors may lack the capacity or knowledge to process the same information effectively (Cho et al., 2013, p. 5). As a result, the impact of ESGD on investment decisions may vary significantly depending on the investor group in question.

The **upcoming study** examines whether ESGD influences a company's ownership structure, with particular attention to institutional investors. Institutional investors, such as pension funds, mutual funds, and insurance companies, play a dominant role in global financial markets and are generally more sensitive to the long-term risks and opportunities associated with ESG factors (Bushee and Noe, 2000, p. 172; Hoq et al., 2010, p. 23). The central research question of this thesis is: Does ESG disclosure have an impact on a company's ownership structure? To explore this question, two key hypotheses will be tested: an introductory hypothesis H1, which posits that ESGD has an impact on ownership structure in general, and the main hypothesis H2, which suggests that ESGD has a positive impact on institutional ownership. Focusing on institutional investors is particularly important given their significant impact on companies and their growing interest in sustainable investing (Caselli et al., 2023, p. 3; Hirst et al., 2023, p. 977). With growing pressure to conform to ESG standards, it could be advantageous for firms to additionally focus on transparency rather than solely engaging in resource-intensive new initiatives. By ensuring transparency about existing ESG practices, companies may attract institutional investors while avoiding the costs, delays, and other hurdles associated with implementing new ESG actions (Cleveland et al., 2023, p. 44 et seq.). The findings of this study can have implications not only for investors but also for

companies seeking to navigate the evolving demands of the capital markets. By understanding how ESGD affects ownership structures, particularly the behavior of institutional investors, companies may better align their reporting practices with the preferences of key investor groups. Moreover, this research contributes to the broader discussion on the role of transparency in corporate governance in an era of increasing sustainability awareness.

The remainder of this thesis is organized as follows. Section 2 examines the theoretical background, focusing on ESG transparency and ownership structure. Each part includes an overview of their economic effects to show the practical implications and emphasize feasible outcomes. The intersection of these two topics and their relationship provides the foundation for the hypothesis development. Section 3 outlines the research design, methodology, and data sources. Section 4 presents the empirical results from the analyses of the relationships between ESGD and ownership structure. Section 5 analyzes the study's results, evaluates the implications of ESGD on ownership structure, and discusses the strengths and limitations of the research. Finally, Section 6 summarizes the key findings and highlights potential avenues for future research, emphasizing the ongoing relevance of ESG disclosure and ownership structure in sustainable investing.

2. Theoretical Background and Hypothesis Development

The following section provides a comprehensive theoretical foundation for the study, focusing on the two fundamental dimensions "ESG transparency" and "ownership structure". Figure 1 illustrates the layout of this section, highlighting how the hypotheses are incorporated within a theoretical framework. Section 2.1 delves into the concept of transparency, its origins, and the evolution of ESG transparency as a new aspect of corporate disclosure. To demonstrate its practical relevance, this section also includes a brief overview of selected economic effects of ESG transparency. Similarly, Section 2.2 explores the concept of ownership structure with particular emphasis on institutional investors and presents a selection of economic implications to highlight its significance. Section 2.3 then explores the relationship between ESG transparency and ownership structure, considering how each factor influences the other. Finally, Section 2.4 narrows the focus to develop the study's hypotheses, explicitly addressing the unidirectional impact of ESGD on ownership structure.

2.1. (ESG-) Transparency

2.1.1. Concept of Transparency

Origins and Basics of Corporate Transparency

Transparency and disclosure are essential concepts to align the interests of companies and their stakeholders, especially between managers and investors. Generally, corporate transparency can be defined as "the availability of

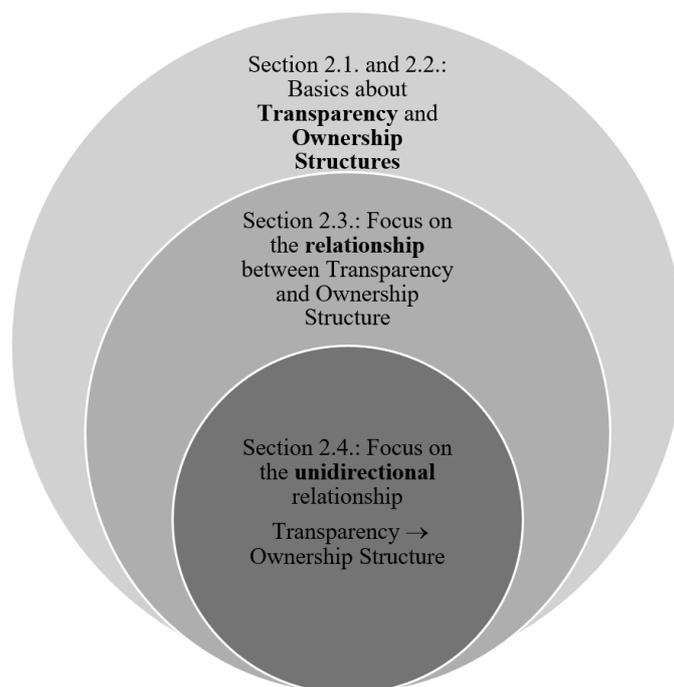


Figure 1: Visualization of the Structure in Section 2

firm-specific information to those outside publicly traded firm“ (Bushman et al., 2004, p. 207). Transparency is essential because managers possess superior insights regarding their firm’s expected future performance, even in efficient capital markets (Healy & Palepu, 2001, p. 420). This informational advantage creates **information asymmetries**, where one party holds more or better information than another. Information asymmetry is a core concept in agency theory that posits an inherent conflict between the interests of a principal (e.g., owners) and the agent (e.g., managers), which can ultimately result in withholding vital information (Jensen & Meckling, 1976, p. 5 et seq.). As Healy and Palepu (2001, p. 420) explain, this agency problem arises as companies and their managers might overstate their value to attract investment. Once the investment is secured, the management may act in ways that expropriate investors’ savings. The resulting information asymmetries, not only between managers and investors but also among various groups of investors (Bishara et al., 2020, p. 1581), place shareholders at a disadvantage, potentially leading to inefficient decision-making or misallocation of resources (Healy & Palepu, 2001, p. 407). Increasing transparency through enhanced disclosure can serve as an effective solution for this problem and reduce information asymmetry (Diamond & Verrecchia, 1991, p. 1325).

Historically, transparency has been predominantly linked to the disclosure of financial information, which serves as a cornerstone of corporate accountability. These disclosure practices are typically categorized into **mandatory** and **voluntary** releases. Mandatory disclosures are enforced by regulatory frameworks, while voluntary disclosures can be

guided – following established standards – or non-guided individual measures, such as conference calls and presentations on websites. A widely recognized framework for **compulsory transparency** is the International Financial Reporting Standards (IFRS), which apply mainly to publicly traded companies (IFRS Foundation, 2024). Adherence to IFRS is mandatory in many countries: Under Regulation (EC) No 1606/2002, all European Union (EU) public companies must prepare their financial statements in compliance with IFRS (European Parliament and Council of the European Union, 2002). In contrast, publicly listed companies in the United States (US) must follow the Generally Accepted Accounting Principles (GAAP), the IFRS-equivalent reporting standard in the US (Code of Federal Regulations, 2023). Moreover, mandatory frameworks on a national level can require additional transparency. For example, the German Handelsgesetzbuch (HGB) regulates compulsory disclosure of both public and private companies based on size and structure (German Commercial Code, 2024). Beyond these mandatory frameworks, companies can adopt **voluntary standards** like those from the International Organization for Standardization (ISO) to enhance their credibility and reputation (International Organization for Standardization, 2024). In addition to adopting these mandatory and voluntary standards, firms can improve transparency through **individual disclosures**, such as management forecasts, analysts’ presentations, conference calls, press releases, websites, and other corporate reports (Healy & Palepu, 2001, p. 406). Lang and Lundholm (1996, p. 468) note that “firms vary substantially in the amount of additional information they provide to the capital markets”, indicating that voluntary

transparency practices can differ widely even among firms operating under the same regulatory environment. This voluntary disclosure can further bridge the information gap between companies and investors as well as among different investor groups. Besides regulatory frameworks, several other factors can influence a company's disclosure practices. Haniffa and Cooke (2002, p. 317) identify critical elements such as the economy, capital markets, enforcement mechanisms, or the cultural context and emphasize that “disclosure practices do not develop in a vacuum, but rather reflect the underlying environmental influences”. Consequently, the effectiveness and nature of disclosures are shaped by a broader set of conditions beyond just compliance.

Although mandatory guidelines and external pressures for disclosure require effort from companies, additional transparency and lower information asymmetries offer substantial **benefits**. For instance, higher transparency can reduce adverse selection costs and increase investors' awareness, which may help attract a larger investor base (Cahan et al., 2016, p. 581; Healy and Palepu, 2001, p. 431). Such disclosures can also reduce estimation risk for investors, thereby supporting a more stable investment environment (Cahan et al., 2016, pp. 581-582). While transparency is primarily advantageous, it may also carry certain **risks**. As H. B. Christensen et al. (2021, p. 1230) point out, “prior literature shows that corporate disclosures can induce proprietary and litigation costs”. Moreover, voluntary disclosure theory suggests that firms with strong economic performance disclose more information (Cao et al., 2019, p. 290). This selective transparency can lead to biases, as recipients may overestimate the company's overall stability and potential.

The increasing focus on transparency has paved the way for a more comprehensive understanding of corporate responsibility. As the corporate governance landscape evolves, investors are placing greater emphasis on non-financial factors in assessing company performance. Consequently, there has been a notable increase in ESGD¹. As stakeholders, including investors, consumers, and regulators, increasingly demand insights into a company's broader impacts², the principles of general transparency provide a vital foundation for understanding the importance of **ESG transparency**. A common definition of ESG disclosure refers to the process of companies reporting on the risks and opportunities related to environmental and social factors, along with the impacts their activities have on people and the environment (e.g. European Commission, 2024). ESGD can be seen as a natural extension of traditional financial transparency, reflecting a more holistic view of a company's performance and respon-

sibilities. Just as financial transparency aligns the interests of managers and investors, ESGD plays a critical role in bridging the information gap regarding a company's societal and environmental responsibilities.

ESG Disclosure as a New Aspect of Corporate Transparency

As transparency widens to include a broader range of corporate impacts, ESG³ Disclosure has emerged as a key extension of corporate reporting. The increase in ESG reports is closely tied to the growing investor interest in non-financial data (Eccles et al., 2011, p. 113). This interest can be explained by the additional informativeness of ESGD, which provides “information about the values of intangible assets – including human capital, natural capital, corporate brands, and general reputation” (Serafeim, 2015, p. 35). ESG reporting covers various long-term, non-monetary, and intangible topics, involving diverse stakeholders and making it distinct from traditional financial reporting (H. B. Christensen et al., 2021, p. 1230).

The growing importance of ESG reporting is also visible in the regulatory landscape, where lawmakers started to integrate environmental disclosure into their guidelines in the last decade. Similar to financial reporting, sustainability reporting can be separated into **mandatory** and **voluntary** disclosures. However, the regulatory landscape for **mandatory sustainability reporting** is relatively new and less established than financial reporting: The EU has led significant developments in establishing and standardizing non-financial reporting. The 2014 Non-Financial Reporting Directive (NFRD) was an early milestone, mandating large companies to disclose key ESG information (European Parliament and Council of the European Union, 2014). As part of the European Green Deal, a more extensive framework for ESGD was introduced with the (EU) 2022/2464 regulation in 2022. This policy, also known as the Corporate Sustainability Reporting Directive (CSRD), will be partly effective from 2024 onward and aims to obligate more companies to provide more detailed sustainability reports (European Parliament and Council of the European Union, 2022), marking a significant advancement in mandatory ESG disclosure across the EU. In contrast, the US has no legal compulsory framework for ESG transparency (DataTracks, 2024). However, the SEC's Regulation S-X, Article 14, will require registrants to provide material climate-related information in their registration statements and annual reports from 2025 onwards (U.S. Securities and Exchange Commission, 2024). In addition to these mandatory frameworks, many companies engage in **voluntary disclosure**. Established frameworks like the Global Reporting Initiative (GRI)

¹ According to the Governance & Accountability Institute (2024), the percentage of S&P 500 companies that published sustainability reports or disclosures increased from 20% in 2011 to 98.6% in 2023

² Surveys show a growing interest among different stakeholder and shareholder groups, e.g. PwC (2023) find an strong increase (+15%) of predominantly institutional investor demand for ESG disclosure between 2021 and 2023. The Morgan Stanley Institute for Sustainable Investing (2024) finds that 65% (66%) of individual investors in America (Europe) increased their interest in sustainable investing between 2022 and 2024.

³ It's important to note that academic literature regularly uses the term Corporate Social Responsibility (CSR) instead of ESG, often even interchangeably. While there are subtle differences between both terms, for the purposes of this thesis, CSR and ESG are considered equivalent concepts and are treated equally. Thus, this thesis only uses the term “CSR” when referring directly to quotes and academic studies. In all other cases, the term ESG is used.

and the Sustainability Accounting Standards Board (SASB) support companies aiming to go beyond compliance and proactively share their sustainability practices. While not legally binding, these standards offer companies guidance to demonstrate a deeper commitment to transparency, often improving their credibility and fostering stakeholder trust (Global Reporting Initiative, 2024; Sustainability Accounting Standards Board, 2024). Even though ESG transparency is increasingly recognized and regulated, its reporting environment remains less mature and standardized than traditional financial reporting.

The growing popularity of ESGD may also be linked to its anticipated **positive impact** on capital markets. As discussed in Section 2.1.1, increased general disclosure lowers estimation risk by reducing information asymmetries. Enhanced ESGD can further diminish these asymmetries and ultimately result in a lower cost of equity (Dhaliwal et al., 2011, p. 90; Ramdhony et al., 2024, p. 525). Additionally, Moss et al. (2024, p. 2) discover that ESGD results in significant responses in stock market prices and trading volumes, indicating that non-financial releases contain market-relevant information for investors. A more detailed overview of the economic implications of ESGD will follow in Section 2.1.2 From a company's perspective, ESGD is a strategic tool for communicating environmental performance: Voluntary disclosure theory indicates a positive association between a company's environmental performance and its level of environmental disclosure (Clarkson et al., 2008, p. 304). This suggests that well-performing firms use ESGD to signal their quality and emphasize their strong performance (Cho et al., 2013, p. 3). In contrast, socio-political theories suggest that underperforming firms employ this transparency to justify poor performance and proactively address stakeholder concerns (Patten, 2002, p. 772). Deegan (2002, pp. 290-291) further explains that ESGD serves as a mechanism for these firms to legitimize their operations, potentially benefiting the company and the broader community. The strategic use of ESGD to communicate or justify corporate performance aligns with legitimacy theory, a foundational framework in ESG research that highlights the legitimizing role of ESGD. A standard definition of **legitimacy theory** is presented by Suchman (1995, p. 574):

„Legitimacy is a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions.“

While legitimacy theory broadly applies to all corporate actions, it seems particularly relevant to ESG practices due to its inherent ties to societal norms and values. Through environmental transparency, companies can demonstrate alignment with societal expectations for sustainable behavior and disclosure, fostering stakeholder acceptance and trust. At its core, legitimacy theory posits that “firms have to conform to societal norms in order to prosper” (de Villiers & van Staden,

2006, pp. 763-764). When companies meet social expectations, they uphold a “social contract” with society, whereas breaches in this contract could threaten their survival (Mathews, 1993, p. 26). This idea positions legitimacy as an essential resource for a company's existence, which must be continually monitored and managed (Dowling & Pfeffer, 1975, p. 125). In a legitimacy crisis, companies face a threat to their survival. However, companies do not necessarily need to adjust their behavior in such crises: they can strategically adapt disclosure to reshape perceptions and expectations, ultimately restoring legitimacy (Dowling & Pfeffer, 1975, pp. 126-127). Strategic disclosure – including the timing, content, and framing of ESG data – can influence stakeholder perceptions, making ESG transparency a powerful tool for maintaining legitimacy (de Villiers & van Staden, 2006, p. 764); (Woodward et al., 2001, p. 362). In contrast, even if companies adapt their actual behavior, those corrective actions might stay unnoticed without adequate transparency.

While disclosure is a critical tool for managing legitimacy, it carries inherent **risks**: Companies may limit certain disclosures if they see them as more harmful than beneficial to their legitimacy (de Villiers & van Staden, 2006, pp. 764-768). Thus, while disclosure remains vital for managing legitimacy, companies must strategically balance transparency with possible reputational harm. Further critiques exist regarding the actual informativeness of ESG transparency. Serafeim (2015, p. 35) argues that “the information provided by most sustainability reports, in addition to being less credible and timely than financial reporting data [...], is not placed in the context of a company's strategy and business model”. He also asserts that investors gain only a few insights into a company's management and its plans for long-term value creation (Serafeim, 2015, p. 35). Additionally, the largely voluntary reporting landscape yields complaints from investors about the lack of comparability and verifiability of information (Bernow et al., 2019, p. 6). This raises important questions about the reliability and relevance of the information disclosed in ESG reports and underlines the importance of regulatory guidance.

Existing literature explores ESG reporting from multiple theoretical perspectives, which can be categorized into **socio-political** and **economic** approaches (Cao et al., 2019, p. 289). The economic perspective focuses on how ESGD connects to financial and market-related factors. Since this view directly aligns with the financial interests of shareholders and their investment decisions, Section 2.1.2 will specifically address the economic effects of ESG transparency. In contrast, the socio-political perspective, which is closely tied to legitimacy theory, emphasizes the role of ESGD in the interaction between companies and their broader environment. This perspective is particularly important given that a large part of a firm's environment consists of its investors, whose values and expectations significantly shape corporate behavior. Sections 2.2 through 2.4 will delve into the socio-political perspective, with Section 2.2 offering an overview of the investor landscape, and Sections 2.3 and 2.4 focusing on the connections between ESGD and investors.

2.1.2. Economic Effects of (ESG-) Transparency

Academic literature explores the economic perspective of ESGD through various approaches. One of these approaches is the relationship between disclosure and a firm's **cost of capital**. This link is especially relevant, as the cost of capital plays a critical role for companies and investors: From a company's perspective, the cost of capital impacts its financing and operational decisions (Dhaliwal et al., 2011, p. 60). For investors, the cost of capital is essential as it serves as a discount rate for evaluating investment opportunities (Berk & DeMarzo, 2023, p. 445 et seq.). By capturing associated risks and reflecting the time value of money, this concept helps investors make informed decisions about resource allocation and potential returns. As Dhaliwal et al. (2011, p. 60) note, there is a "longstanding interest" in understanding how various disclosure forms affect capital costs. The general assumption is that higher disclosure, whether financial or non-financial, is associated with a lower cost of capital: The underlying mechanism suggests that higher disclosure leads to lower information risk, which reduces the risk premium required by investors. A lower risk premium implies a decreased incremental return demanded by investors, ultimately leading to lower cost of capital (Healy & Palepu, 2001, p. 430). Prior research on the relationship between disclosure and the cost of capital focuses mainly on **financial** disclosure: The general academic viewpoint is that financial disclosure is negatively associated with the cost of capital (Dhaliwal et al., 2011, p. 62), indicating that more and better disclosure leads to a lower cost of capital. However, more recent studies focus on **non-financial** information, particularly ESGD. For example, Dhaliwal et al. (2011, p. 79) find a negative association between ESGD and the cost of equity capital for US firms with superior CSR performance, highlighting the economic benefits of transparency in non-financial reporting.

Beyond the impact on the cost of capital, transparency also influences other financial dimensions, such as the **liquidity** of a firm's securities. Diamond and Verrecchia (1991, p. 1326) show that "disclosure improves the future liquidity of a firm's securities". When information asymmetries occur, some investors are less willing to trade due to concerns about trading at unfavorable prices, which leads to illiquidity (Dhaliwal et al., 2011, p. 62). Contrarily, additional disclosure helps to reduce these information gaps. For firms with high levels of disclosure, investors can feel more confident that stock transactions occur at a "fair price," resulting in greater liquidity in the market (Healy & Palepu, 2001, p. 429). The benefits of liquidity are twofold: it ultimately lowers a firm's cost of capital (e.g. Amihud & Mendelson, 1986, p. 224); (Belkhir et al., 2020, p. 1) and can address the preferences of certain investor groups, such as institutional shareholders (see Section 2.2.1).

In addition, academic literature links ESGD to other financial benefits. For instance, both Zhong and Gao (2017, p. 20) as well as Dhaliwal et al. (2011, p. 62) report a positive relationship between ESGD and **firm value**. Furthermore, Zhong and Gao (2017, p. 20) argue that ESGD improves **in-**

vestment efficiency by providing more precise insights into a company's long-term sustainability. Another research perspective examines how ESGD influences **analysts**, who are key users of corporate disclosures and play a critical role in determining the economic effects of transparency. For example, Dhaliwal et al. (2011, p. 90) find that voluntary CSR disclosure is associated with "increased analyst coverage, improved forecast accuracy, and reduced forecast dispersion," particularly in firms with strong CSR performance. This valuation stability indirectly lowers a firm's cost of capital through its positive impact on stock liquidity and the lower risk and uncertainty associated with the forecasts (e.g. Dhaliwal et al., 2006, p. 709); (Gebhardt et al., 2001, p. 146). Furthermore, the reduced forecast dispersion shows that ESGD helps reduce information asymmetries not only between managers and shareholders but also among shareholders – in this case, analysts. Further research highlights additional benefits of ESGD, such as its positive impact on corporate **reputation** (Hasseldine et al., 2005, p. 147). Additionally, Cheng et al. (2014, p. 1) provide evidence that transparency around CSR performance helps reduce **capital constraints**, which they attribute to the reduction of information risk.

Altogether, ESGD is associated with a range of potential benefits across various economic dimensions. However, despite these benefits, research also suggests **negative** effects of ESGD in some instances. First, disclosure is inherently associated with the cost of preparation and dissemination of the information (Verrecchia, 1983, p. 179 et seq.). These **costs** can sometimes outweigh the benefits of additional disclosures, indicating that ESGD may have an unfavorable economic impact. Secondly, PwC's survey of investors reveals that 94% of respondents believe that corporate reporting on sustainability performance contains at least some unsupported claims (PwC, 2023, p. 6). This perception raises concerns about **greenwashing** and leads to decreased accuracy of the disclosed information (PwC, 2023, p. 6). Following the logic of improved analyst accuracy and its positive impact on a firm's economics, reduced precision can lead to negative consequences such as increased uncertainty, heightened risk, and higher costs of capital. Furthermore, Yu et al. (2013) discover "abnormal negative stock price reactions to the disclosure of negative sustainability information", suggesting that higher levels of ESGD do not necessarily improve financial outcomes.

Apart from the existing literature that emphasizes the **economic** implications of ESGD, a substantial body of research also explores its **socio-political** perspective. This perspective provides valuable insights into the consequences of increased ESG transparency, particularly in fostering legitimacy and aligning with shareholder expectations. The following sections focus on this socio-political perspective, examining how different investor types, ownership structures, and ESGD practices shape the relationship between companies and their environment.

2.2. Ownership Structure

2.2.1. Concept of Ownership Structure

Conceptual Foundations of Ownership Structure and Investor Classification

Exploring ownership structure is essential for understanding the socio-political aspects of ESGD, as it reveals corporate dynamics and the relationship between a company and its environment. Ownership structure in publicly traded companies is typically defined by **two primary dimensions**: the level of ownership concentration, which distinguishes between firms controlled by a few major shareholders ("concentrated") versus those held by many smaller ones ("dispersed"/"diffused"), and the type of ownership, such as individual or family investors, institutional investors, or other corporations (Thomsen & Conyon, 2012). In this context, the theory argues that if both managers and investors are driven by value maximization, the specific structure of ownership should have little impact on firm outcomes (Caselli et al., 2023, p. 2; Demsetz, 1983, p. 386). Furthermore, Serafeim (2015, p. 35) suggests that if managers act regardless of their firm's ownership structure, investor characteristics like holding period do not matter for corporate value. However, dynamics like information asymmetries in the principal-agent relationship (Jensen & Meckling, 1976, p. 5 et seq.) contradict those assumptions: generally, ownership structures appear more complex and impactful than these theories suggest.

Due to these complexities, shareholders can be categorized by various **characteristics** influencing their behavior and impact on firms. Caselli et al. (2023, p. 1) note that companies frequently have multiple shareholders, "each with unique incentives, investment horizons, governance preferences, and regulatory constraints". These variations lead to differing costs and benefits of ownership across different shareholder types. Moreover, they find that shareholders' motivations for monitoring and its effectiveness can vary significantly (Caselli et al., 2023, p. 1). Consequently, ownership structure plays a crucial role in shaping corporate governance and, ultimately, influencing the firm's performance. To understand this impact more deeply, academic literature offers various perspectives. For example, Wahl (2006, p. 95) discusses classifications of investors along dimensions such as investment horizon (short-term vs. long-term), risk tolerance (risk-averse vs. risk-tolerant), and preference for stock value versus growth potential. He also highlights a common differentiation based on investor type, differentiating between family, institutional, governmental, and corporate shareholders, each with unique implications for corporate strategy and performance. Similarly, Khlif et al. (2017, p. 379) offer a variation of this categorization, noting differences among institutional, foreign, state, and managerial ownership. They emphasize that these owners vary in power, wealth, competence, and non-ownership ties to the firm. Khlif et al. (2017, p. 379) also introduces ownership concentration with

its characteristics and effects as a critical factor in analyzing another dimension of ownership structure. In general, **Institutional investors** play a particularly significant role among different investor types, as their substantial capital often drives corporate governance practices (Bishara et al., 2020, p. 1581; Hoffmann et al., 2022, p. 198 et seq.). Thus, Section 2.2.1 and the rest of this paper will focus on institutional investors as a critical investor group, examining their motivations and behaviors in greater detail.

The Special Role of Institutional Investors in Shaping Corporate Practices

Institutional investors are professional market participants, typically large legal entities that manage substantial portfolios of assets and invest their or others' funds in stocks and other assets. These investors are distinguished from other investors by their expertise, experience, and ability to assess investment risks, and they include entities such as investment companies, pension funds, banks, insurance companies, foundations, private equity firms, or hedge funds (Hoffmann et al., 2022, p. 198 et seq.).

A critical distinction between institutional investors and other investor types is their **information level** about the companies they invest in (Cho et al., 2013, p. 5). Institutional investors and analysts typically possess more information than other shareholder types. This differentiation results from their superior capacity to gather and process information, often due to their resources and expertise (Cho et al., 2013, p. 5). The higher information level of institutional investors is closely related to their behavior in the investment process: First, institutional investors are motivated to perform in-depth evaluations before possible investments. Given their typically large investment amounts, a thorough analysis is crucial to ensure informed decision-making, demanding extensive and detailed information (Hoq et al., 2010, p. 24). Second, institutional investors monitor the company's management during the investment period (Serafeim, 2015, p. 36). A higher information level facilitates this monitoring role.

Among institutional investors, an important differentiator is their **investment horizon**. While some institutional investors may adopt shorter-term approaches, most institutional investors are associated with long-term strategies (e.g. Bushee & Noe, 2000, p. 172). This distinction significantly influences their behavior, including the types of companies they invest in, the criteria they prioritize, and how they manage their portfolios over time. **Short-term** institutional investors tend to prioritize liquidity in their investment decisions. Given the flexibility needed to trade large positions quickly, these investors favor stocks with high trading volumes (Bushee and Noe, 2000, p. 176; Diamond and Verrecchia, 1991, p. 1327). One way to attract these liquidity-focused investors is by enhancing transparency: As highlighted in Section 2.1.2, increased disclosure can improve stock liquidity, making the firm's stock more attractive to such investors. In contrast, **long-term** institutional investors place

less emphasis on liquidity. Their extended investment horizon and significant financial commitments demand a more thorough analysis and a greater reliance on comprehensive, in-depth information about companies (e.g. Bushee & Noe, 2000, p. 172). In this context, enhanced transparency can also attract long-term investors by providing detailed information. Rather than focusing on quick trading, long-term investors are concerned with the stability and sustainable growth of the companies they invest in. As Bushee and Noe (2000, p. 200) observe, long-term investors are more likely to prioritize factors such as strong management practices, growth potential, and corporate responsibility. By attracting long-term investors, firms can establish a stable ownership base, which offers a strategic advantage by reducing stock price volatility from short-term trading activities (Bushee & Noe, 2000, p. 176).

Because of this predominantly long-term focus, institutional investors often consider factors beyond only short-term financial performance in the investment process: They have an increasing **interest in non-financial criteria**, such as a company's social performance, which can positively impact long-term economic outcomes (Hoq et al., 2010, p. 23). More generally, Qa'dan and Suwaidan (2019, p. 35) note that long-term performance can be enhanced "by good management practices such as CSR action". This suggests that many institutional investors are not solely focused on maximizing immediate financial returns but also on supporting companies with sustainable and socially responsible practices. The shift in priorities is summarized by Hirst et al. (2023, p. 977), as "investors are generally willing to forgo some monetary gains to promote social interests." This theoretical framework is supported by evidence showing that a higher focus on ESG criteria is associated with the interest of institutional investors. For example, Kim et al. (2018, p. 1190) observe that firms with better ESG performance tend to appeal to both institutional and individual investors, particularly long-term and low-stake institutional investors with a focus on sustainable investments. These findings emphasize the growing importance of ESG factors in investment decisions, along with the critical need for transparent and detailed disclosure of these ESG factors. Such disclosures are essential for long-term investors, as they provide the necessary information to assess the ESG impact of companies. Altogether, these insights reveal not only the preferences of institutional investors but also establish another link between ESG transparency and ownership structure.

Besides the unique preferences and characteristics of institutional investors, it's also crucial to understand why companies should care about the **influence** of these investors on a firm's management. Primarily because of their typically larger stakes in a company (e.g. Hoffmann et al., 2022, p. 198); (Hoq et al., 2010, p. 24), institutional investors might have a particular interest in actively guiding and shaping their portfolio companies: Investors with substantial holdings are more directly impacted by the quality of management decisions, as larger shareholders stand to benefit more from effective management and incur greater losses

from poor governance (Caselli et al., 2023, p. 3; Shleifer and Vishny, 1997, p. 754). Thus, the power and impact of large institutional investors on their portfolio companies should not be underestimated and, therefore, need to be considered. These investors often impact corporate decision-making, such as investment-, executive appointment-, and disclosure decisions (Ntim & Soobaroyen, 2013, p. 472), and shape corporate governance structures (Bishara et al., 2020, p. 1581). Companies with long-term institutional backing often enjoy strategic advantages. Caselli et al. (2023, p. 3) highlight that such firms tend to have better profitability, lower risk, and more innovation while investing less and offering higher payout ratios. Section 2.2.2 will explore financial implications like these in more detail. From a management perspective, the presence of long-term institutional investors enhances not only financial performance but also alleviates short-term pressures: The investors' long-term horizon relieves management of the pressure to temporarily cut research and development (R&D) expenses to boost earnings (Bushee, 1998, p. 305; Caselli et al., 2023, p. 3). This allows managers to focus on sustainable growth rather than reacting to short-term market demands and feeling forced to meet profit requirements. Through their monitoring role, institutional investors discourage opportunistic or self-serving behavior by management and thus lead to more effective decision-making (Bishara et al., 2020, p. 1581; Cornett et al., 2007, p. 1773). Serafeim (2015, p. 36) argues that monitoring can prevent shortsighted investment behavior and earnings management, ensuring that managers prioritize corporate performance over personal interests. This is especially important for companies that face a high level of information asymmetry and thus have only a few other means to overcome interest conflicts connected to agency theory (Bishara et al., 2020, p. 1582).

Note: To stay within scope, this paper focuses primarily on institutional ownership. However, it is essential to recognize that institutional investors are just one component of a broader shareholder base. As a complement to other investor types, institutional ownership offers distinct advantages that might differ from those of other investor types. Thus, while benefitting from the advantages of increased institutional ownership, companies might not be able to fully exploit the advantages of other shareholder types. Fully capitalizing on these advantages would require a balanced mix that leverages the strengths of various investor types to maximize overall corporate performance.

2.2.2. Economic Effects of Ownership Structures

As broadly illustrated in Section 2.2.1 with the impact of institutional investors, ownership structure can be a powerful determinant of corporate performance. Even early studies such as Brush et al. (2000) identified ownership structure as a promising area for extended research, suggesting that "ownership matters to the relationship between financial decisions and growth of the firm" (Bishara et al., 2020, p. 1581). Although research on ownership structure and its impact on firm performance has produced mixed results, the growing body of evidence highlights its dynamic role in shaping fi-

nancial outcomes. Ownership structure appears to be not a fixed characteristic but a variable factor that can significantly influence a company's performance. This reinforces the need to investigate ownership structure and its dimensions as a critical element in understanding and improving a firm's financial outcomes.

One such dimension is **institutional ownership**, where the academic literature shows mixed results regarding its impact on firm performance according to Bishara et al. (2020, p. 1581). Despite this inconsistency, there is a noticeable tendency in the research suggesting that institutional ownership positively influences various performance measures. Bishara et al. (2020, p. 1586) own findings support this view, demonstrating a "positive significant effect of investment manager ownership on sales growth in high information asymmetry environments." They attribute this effect to the monitoring functions of institutional investors, which ultimately enhances operating performance. Similarly, Cao et al. (2019, p. 289) summarize the common academic viewpoint, emphasizing that large shareholders can generally have a "significant economic impact on the firm". As institutional investors are often substantial shareholders, these findings further reinforce the role of institutional ownership in shaping firm growth and performance.

While institutional investors play a critical role in shaping corporate governance and performance, the broader shareholder base – including family, individual, and transient investors – also exerts significant influence, each group bringing its own dynamics and effects on firm stability and growth. For example, the literature finds that **family or corporate shareholders** positively influence firm performance through sales growth (Thomsen & Pedersen, 2000, p. 702). Such growth is likely driven by the stability and forward-looking approach that these investors often support. Second, research on **individual shareholders** reveals differentiated impacts based on the informational environment. In low information asymmetry settings, individual ownership positively affects sales growth (Bishara et al., 2020, p. 1586). However, Bishara et al. (2020, p. 1587) also note that individual shareholders tend to avoid investing in companies with high levels of information asymmetry, as these conditions present increased uncertainty and risk. This behavior aligns with the expectation that individual investors, often lacking the resources and access to information that institutional investors possess, prefer more transparent and less risky environments (Bishara et al., 2020, p. 1582). The tendency of individual shareholders to avoid investments at high levels of information asymmetry also illustrates heterogeneous information access among different investor types. Finally, **short-term shareholders**, in literature often classified as "transients", exhibit behavior that can lead to negative long-term outcomes for firms. Serafeim (2015, p. 35) points out that managers in firms with a higher proportion of these investor types are more likely to cut R&D spending to meet short-term earnings targets. This trend highlights how the pressure from short-term investors can lead to decisions prioritizing immediate financial results at the expense of long-term innovation and

growth.

2.3. The Relationship between (ESG-) Transparency and Ownership Structure

Sections 2.1 and 2.2 introduced the concepts of transparency and ownership structure and highlighted critical links between them: their relationship is shaped by several factors, including the interaction with liquidity, the role of information asymmetries, and the rising investor preferences for companies' sustainable practices. Generally, the dynamics between ESG transparency and different ownership structures are interrelated and must be examined from **both directions**.

Different ownership types play a crucial role in shaping a company's commitment to ESG initiatives and vice versa. For example, the ownership base can influence the extent of ESG activities (Barnea & Rubin, 2010, p. 84). Conversely, ESG practices can shape a company's investor structure by attracting groups that prioritize ESG criteria differently and often have preferences "beyond shareholder value maximization" (H. B. Christensen et al., 2021, p. 1178). Dynamics like this underscore the intertwined relationship between ownership structure and a company's commitment to ESG initiatives which is already studied in academic literature from many perspectives. Since ownership structure is closely tied to the concept of information asymmetry through varying information levels of different investor groups, it is essential to examine not only the connection between the shareholder base and ESG practices but also how the disclosure about these practices, ESGD, is associated with different investor types. The **impact of ownership structure on ESG transparency** is becoming increasingly important as academic literature expands beyond traditional financial performance to explore the non-financial effects of ownership structure.

Due to their substantial ownership stakes, **institutional investors** have been theorized to play a vital role in shaping corporate disclosure practices. Several studies offer **theoretical explanations** for how and why institutional ownership might influence a company's approach to ESGD: One common argument is that institutional investors, driven by profit motives, may encourage greater ESGD because of the effect on the reputation and legitimacy of companies in their portfolios. For example, Ntim and Soobaroyen (2013, p. 472) suggest that institutional owners actively lobby managers to project a more socially responsible image to win the support of influential stakeholders, ultimately enhancing the company's profitability. Another explanation for increasing ESG transparency links to short-term institutional investors' demand for stock liquidity. Haniffa and Cooke (2002, p. 329) and Diamond and Verrecchia (1991, p. 1326) hypothesize that short-term institutional investors are incentivized to reduce information asymmetries among other shareholder groups to actively increase stock liquidity. Another theory for the positive impact of institutional ownership on transparency is provided by Ntim and Soobaroyen (2013, p. 472), as long-term institutional investors require more information to conduct their monitoring role effectively. Contrary to these

theoretical expectations, **empirical findings** reveal a negative effect of institutional ownership on ESG transparency. For example, Ntim and Soobaroyen (2013, p. 481) report a weak negative association between institutional ownership and ESGD. They explain this finding to the specific characteristics of their sample from South African companies, where institutional investors often rely on direct monitoring practices rather than public disclosures. Similarly, Qa'dan and Suwaidan (2019, p. 40) identify a significant negative relationship between institutional ownership and ESGD, though this outcome remains unexplained in their study.

Government ownership is expected to have a substantial effect on ESGD, as **theory** suggests that higher public pressure plays a critical role in this relationship: Firms with significant government ownership are subject to "higher expectations and scrutiny from the public" and are therefore expected to engage in greater ESGD to meet these demands (Ramdhony et al., 2024, p. 528). However, this positive expectation might not hold in countries with weaker governance. For example, in environments characterized by high levels of corruption, government ownership might lead to lower ESGD, as strong political connections can reduce the likelihood of enforcement from corrupt regulatory bodies (Ntim & Soobaroyen, 2013, p. 472) and raise incentives to hide inefficiencies or unethical practices. **Empirical evidence** of this relationship remains mixed (Ramdhony et al., 2024, p. 528). On the one hand, studies on Chinese firms demonstrate that close political ties allow government-owned companies to escape regulatory scrutiny, leading to poor ESGD (Jia et al., 2009, p. 562; Ntim and Soobaroyen, 2013, p. 472). On the other hand, studies in countries with lower levels of corruption⁴ and less entangled political ties have found a positive link between government ownership and enhanced disclosure practices. For example, Eng and Mak (2003, p. 340) observe that government ownership in Singapore can lead to higher levels of voluntary disclosure, as it helps to address issues related to moral hazard and agency problems associated with state ownership. Similarly, Tagesson et al. (2009, p. 360) find that Swedish state-owned companies tend to disclose more ESG information "to serve as good examples" for other companies. Ntim and Soobaroyen (2013, p. 472) support these findings and explain that government ownership can increase transparency, as the government's influential role as a stakeholder encourages responsible disclosure practices. These findings suggest that government ownership in well-regulated, low-corruption environments can foster greater transparency and accountability through enhanced disclosure practices.

The impact of **individual investors** – in academic literature mainly comprising managers, directors, founders, families, or similar shareholders – on ESGD is also ambiguous. The **theory** provides possible explanations for both directions: one rationale for a positive relationship is that direc-

tors might encourage greater disclosure to improve a company's reputation (Ramdhony et al., 2024, p. 529). Conversely, individual investors may reduce transparency to capitalize on informational advantages and claim a personal benefit from the resulting insider information (Ramdhony et al., 2024, p. 529). Another reason for individual inside investors to lower ESG transparency might be to avoid costs for additional disclosure, which would decrease profits and, thus, their salary. Prior studies have produced mixed **empirical results**, including positive, negative, and non-significant relationships (Ramdhony et al., 2024, p. 529; Said et al., 2009, p. 217).

Academic literature examines the effects of ownership structure not only regarding different investor types but also regarding the **concentration** of shareholdings. This consideration is especially relevant as this paper focuses on institutional investors, which are often associated with higher overall ownership levels (Diamond and Verrecchia, 1991, p. 1327; Hoq et al., 2010, p. 24). Because of this connection, examining ownership concentration becomes valuable even when the primary focus is on understanding the relationship between ESGD and institutional investors. Once again, **theory** can explain both the positive and negative impacts of ownership concentration on ESGD. A common approach for a negative link is based on information asymmetries. In firms with concentrated ownership, the need for extensive disclosure may be reduced as the major shareholders often have direct access to information and thus less information asymmetries (Ntim and Soobaroyen, 2013, p. 472; Qa'dan and Suwaidan, 2019, p. 34). In contrast, less concentrated ownership requires disclosure to reduce informational asymmetries between numerous shareholders and management (Brammer and Pavelin, 2008, p. 124; Haniffa and Cooke, 2002, p. 328). Another potential factor behind lower ESGD in concentrated ownership firms is the cost effect. Cao et al. (2019, p. 290) suggest that controlling, influential shareholders may avoid CSR disclosures to minimize additional economic costs, especially when they do not see direct benefits from such actions. Finally, from a legitimacy theory perspective, companies with concentrated ownership may feel less pressure to disclose information to ensure public accountability when outsider interest is limited and less powerful (Ntim & Soobaroyen, 2013, p. 472). Other scholars argue that ownership concentration can positively impact disclosure under certain conditions. The predominant argument is based on reputation-building, where controlling shareholders might push for more ESGD to enhance the company's reputation and ultimately increase profitability (Cao et al., 2019, p. 290; Qa'dan and Suwaidan, 2019, p. 34.) **Empirical evidence** on the link between ownership concentration and ESGD is also mixed. Brammer and Pavelin (2008, p. 131) and Cao et al. (2019, p. 288) find that higher ownership concentration is associated with significantly lower ESGD. Roberts (1992, p. 609) also reports a negative association, though not statistically significant. However, other studies report a positive relationship. For example, Haniffa and Cooke (2002, p. 340) find a significant positive relation-

⁴ Corruption Perception Index in 2023 according to Transparency International (2023); China: Score 42 (Rank 76), Singapore: Score 83 (Rank 5), Sweden: Score 82 (Rank 6)

ship between ownership concentration and ESGD. Furthermore, Jiang and Habib (2009, pp. 295-298) suggest that the effect of ownership concentration on ESGD may depend on the type of investor who is holding the concentrated ownership: While institutional investors with their large shareholdings tend to push for lower disclosures as they often monitor companies through other mechanisms, government ownership concentration tends to encourage greater voluntary disclosure to legitimize the company's operations and align with societal expectations. These findings reveal an interplay between ownership types and ownership concentration and underscore the value of including ownership concentration in the analysis, even though the primary focus of this study is on ownership types.

Overall, the findings about the impact of different ownership types on ESGD are mixed and inconclusive. Furthermore, theory offers explanations for both positive and negative effects and lacks a strictly unidirectional logic. This lack of clarity diminishes the practical value of existing research. One potential reason for the inconsistency in the results is that most studies take a unidirectional approach, focusing solely on how ownership influences disclosure. However, as Healy and Palepu (2001, p. 431) note, it is "difficult to infer whether disclosure changes followed or preceded changes in variables of interest." This uncertainty makes it challenging to identify the direction of causality in the relationship between ESGD and ownership structure.

Ramdhony et al. (2024, p. 525) further suggest that focusing solely on the direct impact of ownership on disclosure may introduce endogeneity issues, potentially distorting the results. To address these concerns, they employ a Panel Vector Autoregression (PVAR) analysis and examine the **bi-directional** relationship between ESGD and ownership structure, considering simultaneous effects in both directions. Their findings offer critical insights into this connection: ESGD shows a significant negative reaction to both government ownership and concentrated shareholdings, while director ownership has a negative but insignificant effect on ESGD. Moreover, their analysis uncovers the inverse relationship, revealing that ESGD negatively impacts government ownership. This finding can be attributed to firms in poor governance environments being "less interested in indulging in good CSR practice and reporting" (Ramdhony et al., 2024, p. 537). Similarly, ESGD significantly negatively influences director ownership, suggesting that directors may avoid governance disclosures that could constrain their potential for higher remuneration (Ramdhony et al., 2024, p. 537). Finally, they suggest an insignificant negative effect of ESGD on ownership concentration.

While Ramdhony et al. (2024) focus on the bi-directional relationship, existing literature also explores the **uni-directional impact of ESGD on ownership structure** in detail. This perspective suggests the possibility of reverse causality, where ESGD could actively shape a company's ownership structure. Theoretical explanations and studies on how disclosure shapes the investor base are discussed in Section 2.4 Examining this relationship is particularly promising because it

could offer actionable insights for companies – not only into what effect their disclosures have in general but also into how these disclosures can attract particular investor types. This leads to the research question:

“Does ESG disclosure have an impact on a company's ownership structure?”

Understanding this inverse dynamic could provide valuable guidance for companies looking to strategically manage their shareholder base. There is already substantial evidence that companies are actively trying to shape their ownership structures. For example, Kim et al. (2018, p. 1190) find that a firm can “increase its investor base by adopting proactive environmental strategies”. Similarly, Serafeim (2015, p. 36) notices that many corporate managers are now actively seeking ways to attract investors with longer-term investment horizons, particularly in response to pressures from existing stakeholders and the broader market. This reflects a conscious effort to shape the firm's shareholder composition towards stability and long-term alignment. Further evidence of this trend is offered by a survey by Beyer et al. (2014), which finds that nearly all companies have a preference for long-term investors. Although these insights suggest that companies should emphasize attracting long-term investors, this approach may not apply universally. As Paul Polman, the former CEO of Unilever, clarified, “you need to attract a shareholder base that supports your strategy – not the other way around. We actively seek one that is aligned with our longer-term strategy” (Polman, 2012). This view supports the idea that companies influence their ownership structure for strategic reasons.

These examples show that companies increasingly take proactive steps to manage ownership structures. The ability to attract certain types of investors – particularly those with a long-term, stability-focused investment strategy – has become an essential element of corporate strategy. This raises the question of whether ESGD can be another tool for companies to influence their investor base.

2.4. Hypothesis Development

Building on the ambiguous insights of the previous section, this chapter shifts the focus to the unidirectional impact of ESGD on ownership structure. The interest in understanding how disclosure affects a company's shareholder base has a long history. Already in 2001, Healy and Palepu (2001, p. 411) synthesized longstanding research questions on how investors respond to corporate disclosures and if possible reactions differ between fundamental or supplemental disclosures. Without clear answers to these questions, companies were cautious about additional disclosures in the 2000s (Hoq et al., 2010, p. 24). More recent studies demonstrate that there is still insufficient understanding of how transparency, particularly within the ESG context, affects ownership structure. For example, Moss et al. (2024, p. 4) find that only “a small number of prior studies have examined how ESG information influences different classes of investors”, suggesting

a need for future research to explore the overall demand for ESGD and the reaction of different investor classes (Moss et al., 2024, p. 15). Therefore, the analysis in this paper aims to address this **gap** by investigating the unidirectional influence of ESGD on ownership structures, particularly regarding institutional investors. By doing so, it seeks to enhance the predictability of investor reactions and encourage greater transparency.

Increasing predictability is crucial, particularly given the existing literature's diverse theories and aspects of ownership structure. The lack of a cohesive framework leads to ambiguity in understanding investor behavior. For example, firms that restrict their general disclosure practices may attract shareholders that prefer opacity to hide unethical behavior (Ramdhony et al., 2024, pp. 528-529). In contrast, Moss et al. (2024, p. 1) argue that companies engaging ESGD may attract investors with a "taste" for sustainability, suggesting that some shareholders are drawn explicitly to firms that align with their values. This interest often remains even when a company undertakes actions that don't directly contribute to its cash flow, as certain investors prioritize activities with value beyond financial returns (Friedman & Heinle, 2016, p. 740). Moreover, as investors differ in their capability to use disclosed information, different levels of sophistication among investors can further shape how ESGD impacts ownership structure. Kalay (2015, pp. 1005-1006) highlights that firms with higher levels of disclosure tend to attract more sophisticated investors who can effectively analyze and use the disclosed information. Additionally, Kalay (2015, p. 1005) finds a direct reaction from investors as "firms that initiate (cease) an earnings guidance policy experience an increase (decrease) in the proportion of sophisticated investors".

Considering **specific investor types**, academic literature reveals various findings about reactions to ESGD. For example, de Villiers and van Staden (2006, p. 766) find that directors are more likely to invest if their firm has a low ESGD level to avoid scrutiny of their compensation. Ramdhony et al. (2024, p. 527) support this and explain an implicit mechanism: ESGD might reveal damaging information and thus harm executive compensation. As remuneration for high-ranked employees often consists of stock options, lower compensation would mean receiving fewer stocks, resulting in lower director ownership. Regarding private investors, Moss et al. (2024, p. 2) "do not detect a retail investor response to ESG press releases". This indifference might result from private investors' relatively limited information processing capability compared to more sophisticated investors, suggesting that they may struggle to use additional ESG data (Kalay, 2015, pp. 1005-1006).

Another key factor in understanding ownership structure is the **concentration** of shareholdings. According to Ramdhony et al. (2024, p. 530), "CSR may negatively influence investment by block owners, as CSR reporting imposes economic costs for these controlling shareholders". Even though this perspective might explain why shareholders avoid further investment, it seems more aligned with the reverse re-

lationship, as controlling shareholders actively influence the decision to limit additional CSR reporting in order to avoid increased costs (Cao et al., 2019, p. 290). Another explanation for the negative association between ESGD and ownership concentration is that firms with greater general disclosure increase investors' awareness about their existence, resulting in a broader base of investors and, thus, lower ownership concentration (Dhaliwal et al., 2011, p. 62).

These **varying effects** of ESGD on different aspects of ownership structure highlight the complexity of this relationship. Moreover, some of the theories and explanations in this section align closely with those presented in Section 2.3, which explained a reverse impact of ownership structure on ESGD. While the literature does not reach a clear consensus on how ESGD affects each type of investor, there is evidence that ESGD might have a general impact on ownership structure: While transparency can attract certain investors, it may deter others. Given these insights, this paper proposes the following hypothesis *H1*:

H1: ESGD has an impact on ownership structure in general

To establish *H1*, the previous part discussed the connection between ESGD and different ownership dimensions, excluding the group of **institutional investors**. This exclusion is due to the evidence suggesting a stronger impact of ESGD on this investor group than other ownership types. While the prior section presented mixed outcomes regarding the overall influence of ESGD on ownership structure, the literature on possible effects of disclosure on institutional investors has shown more robust analyses. As this body of research has yielded clearer and more consistent results, it's reasonable to establish a separate hypothesis for the relationship between ESGD and institutional investors.

Theoretically, this connection is based on the frameworks already mentioned in the previous sections. The positive impact of ESGD is often driven by the typically long-term investment horizons of most institutional investors, who value additional transparency for legitimizing disclosures and a clearer common understanding of the firm's sustainability strategy (Milne & Patten, 2002, p. 16). Serafeim (2015, p. 36) argue that lower levels of ESGD deter long-term institutional investors, as the resulting information gap increases monitoring costs. Further support for a positive effect of ESGD on institutional ownership comes from research on disclosure and liquidity: As transparency increases stock liquidity, firms get more attractive to short-term institutional investors who value the ease of trading in liquid markets (Diamond & Verrecchia, 1991, p. 1327); (Healy & Palepu, 2001, p. 429).

Empirical evidence from the literature strongly aligns with these theoretical expectations, reinforcing the view that transparency positively relates to institutional investors. For example, Hoq et al. (2010, p. 22) find that CSR reporting attracts institutional ownership, mainly because institutional investors emphasize how companies manage social issues. Furthermore, Dhaliwal et al. (2011, p. 80) observe that

voluntary CSRD initiation attracts dedicated institutional investors due to their monitoring role, enabling better assessment of firm value and management control (Serafeim, 2015, p. 36). Overall, there is a strong tendency in the literature that (ESG-) disclosure positively affects institutional ownership. A detailed overview of studies about this relationship can be found in Appendix 1.

However, much of the existing literature either inherently focuses on general forms of disclosure or blurs the line to specific ESG-related transparency. As a result, it becomes difficult to draw precise conclusions about how disclosures on ESG activities influence ownership structures. Many studies also reveal methodological **weaknesses**, which reduce their reliability. For example, some studies use an unsuitable variable for ESGD, such as dummy variables for the initiation (Dhaliwal et al., 2011) or increase (Healy et al., 1999) of ESGD. Other studies like Hoq et al. (2010) use old data, which might be inappropriate for a dynamic topic such as ESGD. Of all the studies, Ramdhony et al. (2024) stand out as one of the few analyses to precisely address the research question of this paper, but show some weaknesses: On the one hand, their analysis is limited by some methodological shortcomings. As the study is set in the emerging economy of Mauritius, with only 40 listed companies and 400 firm-year observations, the generalizability of the findings is limited. Furthermore, Ramdhony et al. (2024) only apply a multivariate regression technique, which restricts the quantitative interpretability of the results due to the inherent correlation between the shareholdings of different investor types as dependent variables (Dattalo, 2013, p. 3). Additionally, the ESG variable in their study is based on content analysis, which could introduce distortions of the actual ESG transparency level across different firms. On the other hand, their study focuses primarily on the effects of ESGD on director, government, and block ownership, neglecting the critical influence on institutional ownership. This **literature gap** highlights the need for further investigation and is a crucial foundation for the contributions of this study. Given these insights, this paper proposes the following hypothesis *H2*:

H2: ESGD has a positive impact on institutional ownership

By addressing the theoretical and methodological weaknesses in prior research, this study builds on existing literature and aims to extend the understanding of ESGD's impact on ownership structures. The methodology of the upcoming analysis draws upon a broad range of related studies (see Appendix 1) while seeking to provide a more comprehensive and targeted investigation.

3. Methodology

This section outlines the research design and methodological approach used for this study. Section 3.1 describes the data collection process, focusing on the sources for ESG

data and all ownership structure variables. Section 3.2 provides an overview of the variables used in the analysis. Section 3.3 explains the multivariate and univariate regression techniques employed in the study to test the hypotheses. Furthermore, the incremental development of the main regression model is presented. Section 3.4 addresses the research quality and discusses measures and robustness checks taken to ensure the accuracy and reliability of the estimated coefficients.

3.1. Data Collection

The empirical analysis is based on a **sample** from sustainabilityreportingnavigator.com with 873 companies from 27 countries, primarily in Europe and Northern America. The original data set contains ESG performance and ESG disclosure data between 2016 and 2022, sourced from the LSEG⁵ and Bloomberg⁶ databases. Ownership data was added from Moody's Orbis⁷, and control variables were retrieved from LSEG. The integration of ownership and control data required further cleaning due to missing values, resulting in 4,733 firm-year observations for the main analysis (see Section 4.2.2). The datasets were matched primarily via the International Securities Identification Number (ISIN), a universal identification code for public companies, ensuring consistency across databases (ISIN Organization, 2024). The period between 2016 and 2022 is particularly relevant due to the substantial advancements in ESG reporting, as outlined in Section 2.1.1. To ensure data accuracy, ownership details are current as of February 2024, while ESG performance and disclosure metrics include updates from August 2024.

Nevertheless, the data has some **weaknesses**: One limitation arises from potential data distortions in the Orbis database, which lists all publicly available ownership links⁸. This can lead to multiple counts of the same shareholding in complex ownership structures, resulting in an overestimation of ownership. For example, in cross- or chain-ownership cases, where Company A owns 100% of Company B and Company B owns 100% of Company C, both A and B might be listed as 100% shareholders in C, thus overestimating ownership (Bureau van Dijk, 2018, p. 810 et seq.). Another potential distortion involves differences in the timing of ESG and ownership data points. ESG data reflects the respective ratings during a period, typically over the financial year. In

⁵ LSEG (formerly Refinitiv) is a leading provider of global financial data and infrastructure, offering a wide range of data products such as market pricing, economic indicators, risk data and company data, which includes ESG information. (London Stock Exchange Group, 2024)

⁶ Bloomberg provides real-time financial data, news, and analytics, delivering comprehensive market information, pricing, economic indicators, and corporate data, which includes ESG information. (Bloomberg, 2024)

⁷ Moody's Orbis is a comprehensive global database with in-depth information on financials, ownership structures, credit ratings, industry classifications. (Moody's Analytics, 2024)

⁸ The information about ownership links in Orbis is sourced from various channels, including direct company disclosures (such as annual reports, shareholder lists, and subsidiaries), official bodies, information providers, SEC filings, stock exchanges, and direct correspondence with companies. (Bureau van Dijk, 2018, p. 830 et seq.)

contrast, ownership data refers to the latest published shareholding information at a specific point in time within that same financial year. This discrepancy could lead to misalignment between the datasets.

Additionally, this thesis has incorporated **Artificial Intelligence**, specifically ChatGPT, to assist in the coding and wording processes throughout the writing.

3.2. Variables

The main explanatory variables for testing the hypotheses from Section 2.4 are the **ESGD Score (ESGD)** and **ESG Performance Score (ESG)**. Using both measures allows for a clear separation of disclosure effects from actual ESG performance, helping to isolate the influence of ESG transparency on ownership structure (H. B. Christensen et al., 2021, p. 1231). The ESGD Score is obtained from Refinitiv ("CSR Sustainability Reporting Score") and reflects the company's transparency regarding its sustainability practices. The ESG Performance Score, also from Refinitiv, measures environmental, social, and governance performance. Both scores are relative as they measure a company's performance compared to its peers within the same industry. For further analyses and robustness checks, the Bloomberg equivalents as absolute measures of ESGD and ESG Performance are used in equivalent regression models. Unlike most existing studies, the ESGD measures from Bloomberg and Refinitiv prioritize overall ESG transparency: by concentrating on general ESG disclosure, these measures avoid the distortions that can arise from over-specifying analyses such as the focus on ESG press releases in the study by Moss et al. (2024).

Additionally, an interaction term between the ESGD Score and the ESG Performance Score ($ESGD * ESG$) is included in most regression models from Section 3.3. This term captures the notion that higher levels of disclosure could have a more substantial impact when the company's disclosed performance values are strong. This idea aligns with Dhaliwal et al. (2011, p. 80), who find a significant positive association between ESGD and institutional ownership for firms with superior ESG performance. In contrast, their findings for companies with lower ESG performance were insignificant. The **ownership data (PctOwn)** from Moody's Orbis database comprises all publicly available ownership links. Thus, it does not account for all shareholder types and notably excludes private and retail investors who do not disclose their shareholdings to the public. As a result, only an average of 52.37% of total ownership is captured, as described in Appendix 3. To enable a more detailed analysis that fits the existing literature's findings, the ownership data is transformed and aggregated into four main investor types: institutional investors ($PctOwnInst$), individual investors ($PctOwnInd$), government ($PctOwnGov$), and corporations ($PctOwnCorp$).

Derived from standard disclosure literature, various **control variables** are included in the analysis to account for firm characteristics and market conditions that might influence ownership besides ESGD and ESG performance. Firm size is measured by the logarithm (\log) of total reported assets

($\log(totAsts)$). Risk is captured by two dimensions: market beta ($Beta$) and leverage (Lev). The beta values are computed over different time horizons based on data availability. Leverage is the ratio of total debt to total capital, reflecting a company's financial risk. Profitability is measured by earnings per share (EPS), and growth is represented by the 3-year compound annual growth rate ($CAGR$) of revenues ($Grwth$). The liquidity of the stocks is captured by the logarithm of the average daily traded value over the past year ($TrdVol$), indicating how easily a company's shares can be bought and sold. Finally, firm performance is measured through two variables: return on assets (ROA) and the market-to-book ratio (MTB), offering insights into operational efficiency and market valuation.

To control for unobservable differences across industries ($Industry$), countries ($Country$), and years ($Year$), the regressions include **fixed effects** for these dimensions. In a variation of the main regression model, industry- and country-fixed effects are replaced by firm-fixed effects to account for additional firm-level characteristics that may influence the observed relationships. Appendix 2 provides a detailed overview of all variables used in the empirical analyses.

3.3. Regression Technique

The analysis employs two similar regressions to test the hypotheses: Both estimate the percentage of ownership ($PctOwn$ and $PctOwnInst$, respectively) as a function of the Refinitiv measures for $ESGD$ and ESG , their interaction term $ESGD * ESG$, as well as various control variables and fixed effects to control for other possible influences on ownership structure. Both models only differ in their treatment of ownership as the dependent variable. For $H1$, a **multivariate model** (Härdle et al., 2024, p. 443 et seq.) estimates the effects of ESGD and other explanatory variables on a set of dependent variables $PctOwn$ (representing corporate, government, individual, and institutional ownership) simultaneously. The multivariate approach allows the analysis of impacts on multiple ownership types at once, reflecting the natural interdependencies and correlations among the complementary shareholder groups. The regression model is the following:

$$PctOwn_{i,t} = \alpha + \beta_1 ESGD_{i,t} + \beta_2 ESG_{i,t} + \beta_3 ESGD_{i,t} * ESG_{i,t} + \beta_4 \log(TotAsts_{i,t}) + \beta_5 Beta_{i,t} + \beta_6 Lev_{i,t} + \beta_7 EPS_{i,t} + \beta_8 Grwth_{i,t} + \beta_9 TrdVol_{i,t} + \beta_{10} ROA_{i,t} + \beta_{11} MTB_{i,t} + Year_{i,t} + Industry_{i,t} + Country_{i,t} + \varepsilon_{i,t}$$

The results for the multivariate model are shown in Table 3 in Section 4.2.1 In contrast, a **univariate Ordinary Least Squares (OLS) regression model** for $H2$ estimates the impact on the percentage of institutional ownership $PctOwnInst$ as the only dependent variable. The model

is built **incrementally** to address the varying approaches and methodologies used in similar studies: the initial regression model includes *ESGD* as well as country-, industry- and year-fixed effects (Model I), followed by the addition of *ESG* (Model II). The interaction term *ESGD * ESG* is then incorporated into a separate model to test whether higher levels of ESG performance amplify the effects of disclosure (Model III). Finally, control variables are added to account for firm-specific characteristics. A comparison is made between the full model (Model V) and a simplified version that excludes the interaction term (Model IV). The complete model, hereafter referred to as the “**main model**” or “**Model (V)**”, can be expressed as follows:

$$\begin{aligned}
 PctOwnInst_{i,t} = & \alpha + \beta_1 ESGD_{i,t} + \beta_2 ESG_{i,t} + \\
 & \beta_3 ESGD_{i,t} * ESG_{i,t} + \beta_4 \log(TotAst_{i,t}) + \\
 & \beta_5 Beta_{i,t} + \beta_6 Lev_{i,t} + \beta_7 EPS_{i,t} + \\
 & \beta_8 Grwth_{i,t} + \beta_9 TrdVol_{i,t} + \beta_{10} ROA_{i,t} + \\
 & \beta_{11} MTB_{i,t} + Year_t + Industry_i + \\
 & Country_i + \varepsilon_{i,t}
 \end{aligned}$$

To include another common approach in similar studies such as Hoq et al. (2010) or Serafeim (2015), two further models replace the fixed effects for industry and country with firm-fixed effects to capture more nuanced firm-specific influences (Model VI and VII). However, using firm-fixed effects may not be entirely suitable given that the sample has a short period of six years, leading to a maximum of six observations per company. Additionally, the *ESGD* values within a company show little variance⁹ over such a short time frame, particularly when considering relative changes, which can distort the estimation of the coefficients in models with firm-fixed effects (Plümper & Troeger, 2019). Figure 2 provides an overview of the incremental steps and the respective regression models. The results for Model (V) and its variations (I), (II), (III), (IV), (VI), and (VII) are shown in Table 4 in Section 4.2.2. The same procedure was repeated using absolute *ESGD* and *ESG* measures from Bloomberg, with results presented in Appendix 7.

To gain a more comprehensive understanding of the relationship between *ESGD* and ownership dynamics, further analyses are built upon Model (V). These models examine the research question from slightly different perspectives, extending the main analysis with different variations.

One critical consideration for this analysis is the **time shift** in investor reactions to *ESGD*. Ramdhony et al. (2024, p. 528) emphasize that CSR reporting typically does not lead to immediate rewards, as investors naturally react over time. Such reactions are often delayed, as investors need time to assess the disclosed information before adjusting their ownership stakes. Therefore, accounting for this time gap in the regression model is crucial to ensure the model's validity in

accurately testing *H2*. The main model already integrates a lag through the inherent definition of the variables: As explained in Section 3.1, the inconsistent timing results as *ESG* ratings reflect a whole period, while ownership data captures a specific point in time. However, the length of this implied lag is not standardized across companies, introducing uncertainty into the analysis. To address this, a variation of Model (V) tests for delayed investor reactions by estimating *PctOwnInst* between $t + 1$ and $t + 5$ relative to *ESGD* at $t = 0$. This allows for a more detailed understanding of how the timing of disclosures affects ownership changes. Given the ambiguous theories and findings in existing research, additional adaptations of Model (V) are conducted to explore patterns and relationships more thoroughly. One variation involves modeling **changes** in *ESGD* ($\Delta ESGD$) against changes in institutional ownership ($\Delta PctOwnInst$). This approach seeks to determine if adjustments in disclosure practices lead to portfolio adjustments of institutional investors. Model (V) is further applied to **subgroups** based on the bottom and top quartiles of stock liquidity to examine differences in the association between *ESGD* and institutional ownership. This approach is motivated by the idea that institutional investors, especially those with a short-term horizon, are particularly interested in companies with high stock liquidity. Moreover, a **PVAR model** is employed as an alternative to traditional linear regression to address potential endogeneity issues and reduce the risk of biased estimates. This model, as suggested by Ramdhony et al. (2024, p. 526) and Jouda (2019, pp. 301-302), captures dynamic relationships between variables and thereby addresses potential issues of simultaneity and reversed causality. Additionally, a **Difference-in-Difference approach** is applied to compare changes in institutional ownership between two groups: a treatment group consisting of companies with a significant increase in *ESGD* during the observation period and a control group of companies with stable *ESGD* levels. To add another dimension of ownership discussed in the literature, an alternative model includes **ownership concentration** as the dependent variable. This approach provides a way to examine another characteristic of institutional ownership, as these investor groups are frequently associated with higher total ownership levels and, thus, with eventually more concentrated ownership. An overview of all additional regression models, their formulas, and their results is shown in Appendix 8. The main results are summarized in Section 4.2.3.

3.4. Research Quality

Given the inconsistent results in existing literature and concerns about endogeneity, it is essential to confirm the robustness and credibility of Model (V). Therefore, a selection of common robustness checks should assess validity, reliability, and objectivity as the fundamental quality criteria in quantitative research (O'Dwyer & Bernauer, 2014, p. 277 et seq.). Additionally, these tests will address key OLS assumptions, including linearity, exogeneity, homoscedasticity, independence, and normality of residuals (Wooldridge, 2020, p. 117 et seq.). This approach ensures that the *ESGD* coefficient

⁹ Mean standard deviation of *ESGD* within companies between 2016 and 2022: 5.6; Median value: 2.1

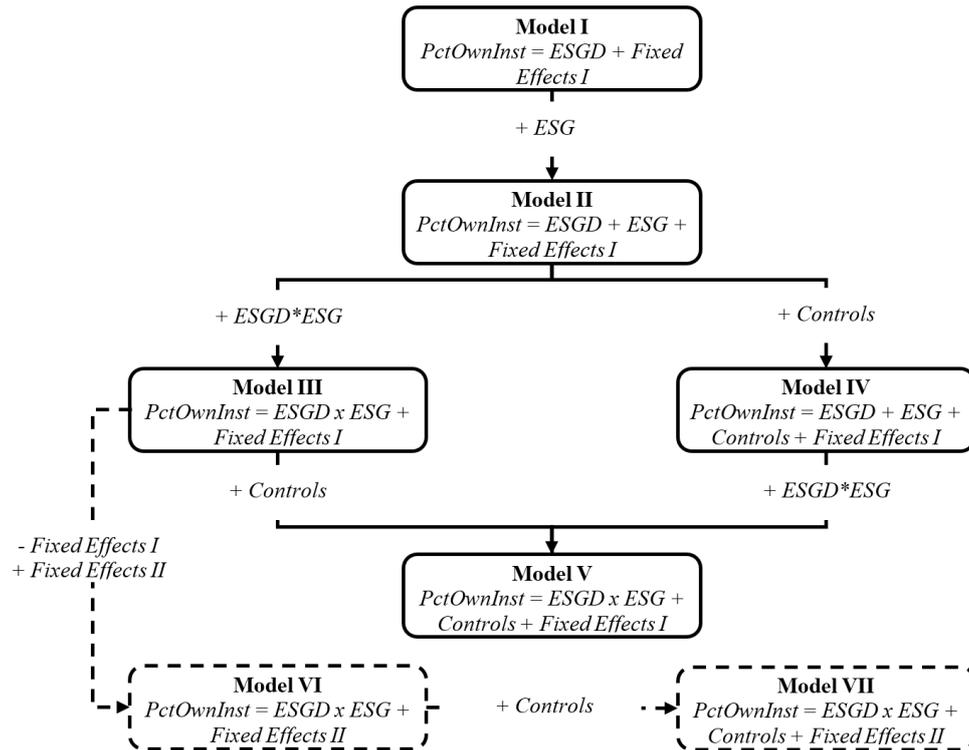


Figure 2: Incremental Development of the Main Regression Model

Illustration of the stepwise development of the univariate regression models to analyze the association between “ESGD” and “PctOwnInst”. **Model I** examines the isolated impact of ESGD on institutional ownership, only controlling for Fixed Effects I (country-, industry-, and year-fixed effects). Each subsequent model builds upon this foundation by incorporating additional variables and interactions. The main model, **Model V**, is the culmination of this development and serves as the basis for all further analyses. To ensure robustness, **Model VI** and **Model VII** replace Fixed Effects I with Fixed Effects II by switching from country- and industry-fixed effects to firm-fixed effects.

is accurately estimated and its suggested effects are isolated from other influencing variables. This separation is especially relevant in ESGD research, where “disclosure changes [...] are likely to coincide with changes in firm economics and governance” (Healy & Palepu, 2001, pp. 430-431). Such corporate shifts might directly imply changes in ownership structure, making it vital to isolate the effects of ESGD from additional control variables.

To ensure the **validity** of the model, several steps are taken. The incremental development of the main regression model by adding explanatory variables and incorporating fixed effects supports the exogeneity. This approach controls for unobserved factors and reduces omitted variable bias. Furthermore, a regression model without outliers is calculated to maintain the accuracy of the estimated relationships, thereby supporting internal validity. Outliers regarding the dependent and the independent variables are identified using Cook’s distance (with a threshold of $>4/n$) and removed from the estimation process to prevent distortions of the results. More support for validity is provided by a Lasso regression, which is employed to refine the model by identifying and eliminating unnecessary variables, thereby enhancing interpretability and reducing omitted variable bias. This helps to ensure that the model specification aligns closely

with the theoretical constructs of interest. Furthermore, a Box-Cox transformation is applied to determine the optimal transformation for the dependent variable, improving the linearity and residuals’ normality. A log transformation for the leading independent and dependent variables further addresses the linearity assumption of OLS.

To reinforce the **reliability** of the findings, further checks of the main regression model are conducted. The Variance Inflation Factor (VIF) values are calculated to assess multicollinearity among the independent variables. Additionally, Bootstrapping is applied to evaluate the robustness of the estimated coefficients by constructing confidence intervals for the estimations. This method provides insights into the precision of the model. The Durbin-Watson test is employed to assess the independence of errors by testing for autocorrelation in the residuals. This check is critical for ensuring the OLS assumption that the residuals are uncorrelated across observations. The Breusch-Pagan test examines homoscedasticity across all levels of the independent variables. This test addresses the OLS assumption of constant error term variance, making it an essential check for reliability. To further account for potential heteroscedasticity, robust standard errors were used in another coefficient test. This adjustment provides more reliable inference from the model estimates,

allowing for unbiased interpretation of coefficients.

Automated techniques like Lasso regression and bootstrapping ensure **objectivity** and minimize biases related to model assumptions.

An overview of all robustness tests and their results is shown in Appendix 15. These comprehensive robustness checks and their consistency with the main regression results help to confirm the credibility of the findings, ensuring that the impact of ESGD is accurately estimated.

4. Empirical Results

The following section presents the findings of the empirical analysis conducted in this study. Section 4.1 provides descriptive statistics and correlation analyses of the variables used in the regression models, offering an initial overview of the data and relationships between key variables. Section 4.2 summarizes the regression results, divided into three parts. Section 4.2.1 presents the findings from the multivariate analysis, examining the impact of ESGD on various ownership types simultaneously. Section 4.2.2 focuses on the univariate regression results, explicitly addressing the relationship between ESGD and institutional ownership. Section 4.2.3 covers the findings of additional analyses to further explore and validate the results of Section 4.2.2.

4.1. Descriptive Statistics and Correlation Analysis

The **descriptive statistics** in Table 1 provide an overview of the variables used in Model (V). The mean ESGD score of 58.8 suggests that the companies in the sample exhibit slightly higher transparency on average than their peers. Furthermore, they tend to act more ESG-friendly than comparable organizations due to their average relative ESG Performance score of 66.72. The mean beta value of 0.97 indicates that the average systematic risk of the companies in the sample is aligned with the overall market. The leverage ratio in the data sample has a mean of 40.74%, implying a moderate level of leverage across the companies. Institutional investors have an average share of 33.25%. The maximum value of 253.59% can be explained by the limitations of the Orbis dataset, as potential multiple counting of shareholdings can result in values above the natural limit of 100% (see Section 3.1).

Figure 3 illustrates the recent **development** of ESGD and the ownership distribution across different shareholder groups. Over the time horizon of this study, ESG transparency was continuously rising, resulting in an overall increase in the average ESGD level of 18.43% between 2016 and 2022. While average ownership of individuals, corporations, and government remains almost constant, institutional ownership declined from 2018 to 2022. This indicates a shift where institutional investors divested from the sample companies, transferring respective shares to "free float" owners like private investors who are not captured by the Orbis data set.

Figure 4 offers additional insights into the industries and countries in the sample and the respective levels of

ESGD and institutional ownership. No discernible **pattern** emerges for ESGD, which remains consistent across sectors and regions. However, Western Europe stands out for having slightly lower ESGD scores across most industries despite the more advanced and often mandatory ESGD standards in developed Western European countries (see Section 2.1.1). In contrast, institutional ownership exhibits a clear geographic pattern: higher institutional shares in the sample companies are observed in Northern Europe and Northern America. At the same time, Western and Southern Europe show lower levels of institutional investment across industries. Overall, companies in the sample are primarily headquartered in Northern and Western Europe as well as in North America, as indicated by the high numbers of data points.

Table 2 **displays** the correlations between the explanatory variables and helps to identify potential multicollinearity issues in the dataset. The data show moderate correlations, with the highest absolute value of 0.45. These relatively low correlations, combined with the low VIF values shown in Appendix 20, minimize immediate multicollinearity concerns in the database. A positive correlation of 0.32 is observed between *ESGD* and *ESG*, which is expected since ESGD is a component of the overall ESG score and directly influences it. Academic literature also explains this relationship through voluntary disclosure theory, as companies with better ESG performance tend to disclose more information (see Section 2.1.1). In contrast, companies with poor ESG performance appear more hesitant to disclose this information. With a correlation coefficient of 0.45 between *ESG* and *TotAsts*, larger companies tend to outperform smaller peers in ESG practices. Additionally, a 0.36 correlation coefficient between *TotAsts* and *Lev* indicates that larger firms rely more on debt capital. A similar relationship exists between *TotAsts* and *TrdVol*, with a correlation of 0.38. Lastly, a moderate positive correlation of 0.38 between *MTB* and *ROA* exists. This is reasonable since the regression model uses both variables as controls for firm performance from different perspectives.

4.2. Regression Results

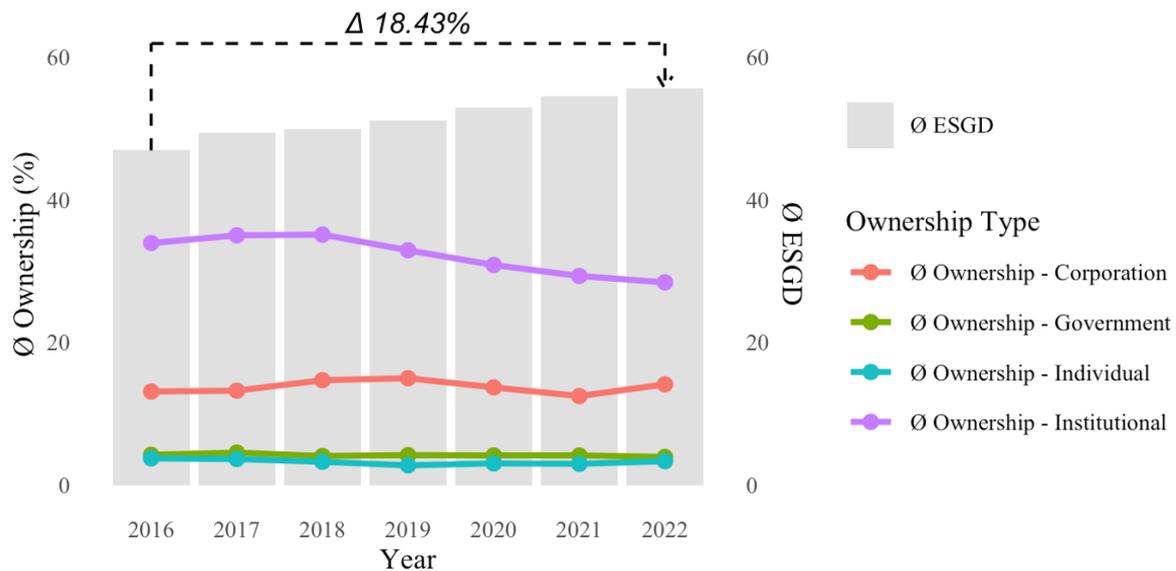
4.2.1. Regression Results: Multivariate Analysis

Table 3 presents the results of the **multivariate regression** analysis, estimating the effects of *ESGD* and *ESG* on multiple ownership types (corporate, government, individual, and institutional ownership) simultaneously. The ESGD and ESG scores used in the analysis are derived from Refinitiv and represent relative measures. Given the complexity of the multivariate regression and hypothesis *H1* focusing on the impact of ESGD on ownership structure overall, the analysis is limited to a qualitative summary of the direction and significance of possible effects. The findings indicate the significance of nearly all variables in the model. Notably, *ESGD*, *ESG*, and *ESG * ESGD* all suggest highly significant effects at the 0.1% level. The same model with Bloomberg's absolute equivalents is included in Appendix 4 and reveals some differences: While *ESGD* remains significant, the coefficient

Table 1: Descriptive Statistics

	<i>N</i>	<i>Mean</i>	<i>Median</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
<i>Refinitiv ESGD</i>	4,733	58.80	57.76	14.68	0	90.80
<i>Refinitiv ESG</i>	4,733	66.72	69.85	16.44	3.91	95.74
<i>TotAsts (mio)</i>	4,733	133,897	16,218	436,815	56	7,967,699
<i>Beta</i>	4,733	0.97	0.94	0.44	-0.45	3.31
<i>Lev</i>	4,733	40.74	40.04	21.24	0	99.49
<i>EPS</i>	4,733	6.30	2.30	39.29	-257.20	1,595
<i>Grwth</i>	4,733	7.39	4.87	20.17	-80.83	613.63
<i>TrdVol (mio)</i>	4,733	817	68	2,632	0	30,172
<i>ROA</i>	4,733	7.52	5.87	12.35	-38.18	292.58
<i>MTB</i>	4,733	5.37	2.34	22.80	0.16	801.50
<i>PctOwnInst</i>	4,733	33.25	29.08	20.51	0	253.59

Summary statistics for the dataset after the data cleaning process, covering all observations and variables included in Model (V). TotAsts and TrdVol are transformed and presented in millions (mio). All variables are defined in Appendix 2. Summary statistics for all variables and the whole sample before cleaning can be found in Appendix 3.

**Figure 3:** Development of Ownership and ESGD between 2016 and 2022

Development of ownership and ESGD over time. Ownership is aggregated as average ownership percentages for the different investor groups for each year. ESGD comprises the average absolute Bloomberg ESGD score for all companies annually (from 0 to 100). The absolute Bloomberg measure was selected to provide actual absolute levels and prevent distortions of the relative measure due to balancing effects within individual countries and industries.

is weaker with significance only at the 5% level. *ESG* remains significant at the 0.1% level, but *ESGD * ESG* shows no significant connection. Furthermore, the equivalent with firm-fixed effects instead of country- and industry-fixed effects results in significant estimations (at 0.1% level) for the relative and absolute versions (see Appendix 5 and Appendix 6).

4.2.2. Regression Results: Univariate Analysis

Table 4 presents the results of the **univariate regression** analysis testing *H2*, which estimates the effect of ESGD on institutional ownership specifically. As explained in Section 3.3, this analysis is based on Model (V) and several variations that introduce incremental adjustments.

Furthermore, this approach includes the **relative ESG measures** from Refinitiv for disclosure and performance scores. For *ESGD*, Model (V) reveals a positive coefficient of

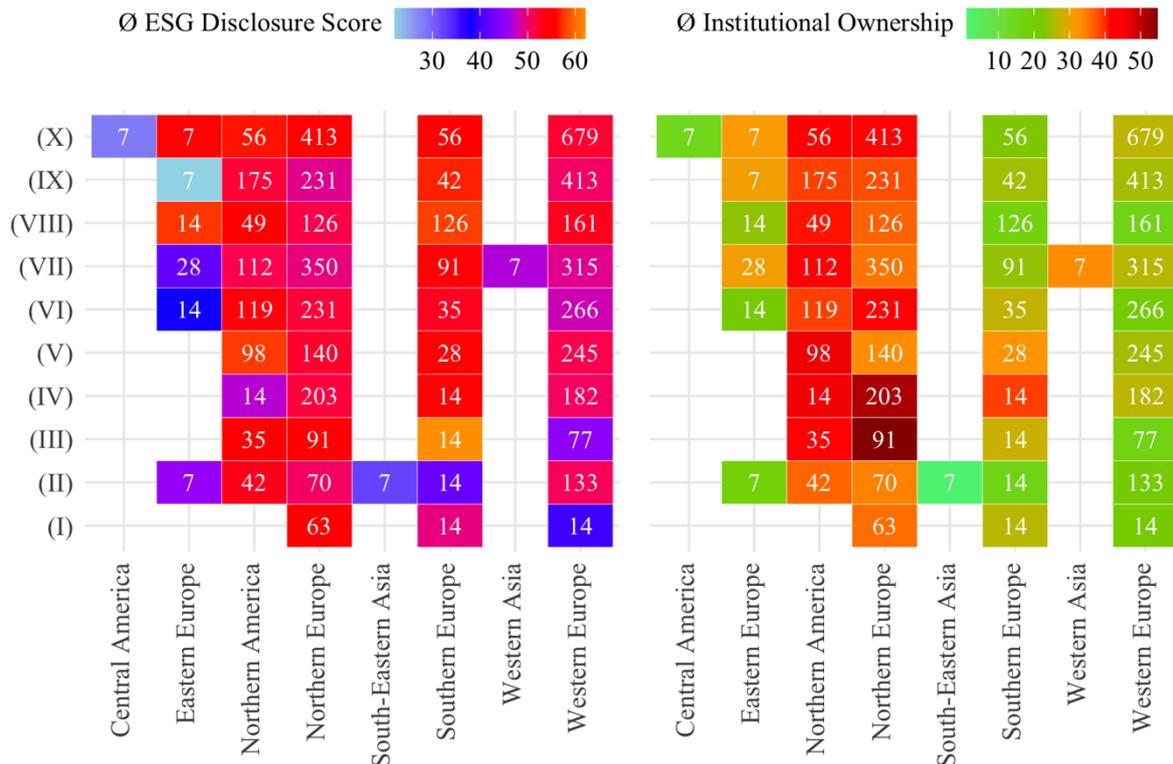


Figure 4: Region-Industry-Matrix for ESGD and Institutional Ownership

Region-Industry-Matrix that visualizes the average Bloomberg ESGD score on the left and the average institutional ownership on the right for all Region-Industry combinations. The Bloomberg measure was selected to provide actual absolute levels and prevent distortions of the relative measure due to balancing effects within individual countries and industries. This aggregation of countries and industries from more detailed values serves two purposes: a) for illustration and b) to identify more general and reliable patterns, as the original dataset contains limited observations for some combinations. The Roman numbers assigned to each industry are as follows: (I) Containers & Packaging, (II) Transportation & Logistics, (III) Hospitality & Entertainment, (IV) Real Estate & Construction, (V) Healthcare & Pharmaceuticals, (VI) Consumer Goods & Retail, (VII) Financials, (VIII) Energy & Utilities, (IX) Technology, Media, and Telecommunications (TMT), and (X) Industrial & Manufacturing. The numbers within each field represent the number of observations included in that specific combination, while white fields indicate combinations with no observations.

Table 2: Correlation Matrix

	ESGD	ESG	TotAsts	Beta	Lev	EPS	Grwth	TrdVol	ROA	MTB
ESGD	1.00									
ESG	0.32	1.00								
TotAsts	0.22	0.45	1.00							
Beta	0.05	0.12	0.18	1.00						
Lev	0.13	0.20	0.36	0.08	1.00					
EPS	0	0	0.10	0.02	-0.03	1.00				
Grwth	-0.12	-0.20	-0.12	-0.09	-0.13	0.02	1.00			
TrdVol	0.23	0.27	0.38	0.11	0.07	0.03	-0.04	1.00		
ROA	0.05	-0.08	-0.26	-0.09	-0.24	0.06	0.10	0.09	1.00	
MTB	0.08	-0.03	-0.09	-0.02	0.11	0	0.06	0.06	0.38	1.00

Matrix for Pearson correlations of the independent variables in the main model. All variables are defined in Appendix 2.

Table 3: Multivariate Regression Results

	<i>DF</i>	<i>Pillai</i>	<i>approx F</i>	<i>num Df</i>	<i>den Df</i>	<i>Pr(>F)</i>	
<i>Refinitiv ESGD</i>	1	0.055	67.193	4	4636	0.000	***
<i>Refinitiv ESG</i>	1	0.026	31.271	4	4636	0.000	***
<i>Refinitiv ESGD*ESG</i>	1	0.006	7.382	4	4636	0.000	***
<i>Log(TotAsts)</i>	1	0.067	82.773	4	4636	0.000	***
<i>Beta</i>	1	0.004	4.919	4	4636	0.001	***
<i>Lev</i>	1	0.011	12.781	4	4636	0.000	***
<i>EPS</i>	1	0.008	9.383	4	4636	0.000	***
<i>Grwth</i>	1	0.002	2.543	4	4636	0.038	*
<i>Log(TRVOL)</i>	1	0.369	678.653	4	4636	0.000	***
<i>ROA</i>	1	0.007	8.238	4	4636	0.000	***
<i>MTB</i>	1	0.001	1.675	4	4636	0.153	
<i>Fixed effect: Year</i>	6	0.034	6.571	24	18556	0.000	***
<i>Fixed effect: Industry</i>	52	0.284	6.822	208	18556	0.000	***
<i>Fixed effect: Country</i>	24	0.406	21.833	96	18556	0.000	***

, **, *** Indicate statistical significance at the 5%, 1%, and 0.1% level, respectively

This table presents the results of the multivariate regression analysis (multivariate analysis of variance; „MANOVA“) on *PctOwnInst*, *PctOwnCorp*, *PctOwnGov* and *PctOwnInd*, as introduced in Section 3.3. All variables are defined in Appendix 2.

0.072, which is statistically significant at the 5% level. This finding indicates that a one-unit increase in the ESGD score is associated with a 0.072 percentage point increase in institutional ownership. Furthermore, the analysis demonstrates a positive association between *ESGD* and *PctOwnInst* across six of the seven model variations, with four suggesting statistically significant relationships. Additionally, institutional ownership is found to mostly react significantly positively to *ESG*, with a coefficient of 0.242 in Model (V). This positive significance is consistent across all models that include country, year, and industry as fixed effects. However, a negative response is observed in the models that incorporate firm-fixed effects instead. The mainly positive coefficients of *ESGD* and *ESG* are at least partially offset by their interaction term *ESGD * ESG*, which is significantly negative in Model (V). The adjusted R-squared values range from 0.460 in the simplest Model (I), which includes only *ESGD* and fixed effects, to 0.826 in Model (VII) with firm fixed effects and all control variables included.

The equivalent analysis with **absolute ESG measures** from Bloomberg is presented in Appendix 7. Table 5 compares the coefficients for the main variables *ESGD*, *ESG*, and *ESGD * ESG* between Bloomberg's absolute measures and Refinitiv's relative measures across all regression models (I) to (VII). In Model (V), the coefficients exhibit notable differences for the absolute Bloomberg measures: Contrary to the anticipated positive association, the *ESGD* variable reveals a non-significant negative impact on institutional ownership. *ESG* demonstrates a positive effect that lacks statistical significance, in contrast to the significant findings noted ear-

lier for the Refinitiv measure. The interaction term, though positive, does not reach significance. Overall, the effects derived from the absolute ESG measures are less consistent and weaker across the various models than those observed with relative ESG data.

Due to the heterogeneous results obtained from the Bloomberg measures and to maintain the study's scope, the focus of further analyses will be on Refinitiv measures. This decision is driven by the need for consistency and clarity, given the robust findings of the Refinitiv data. Focusing on these relative measures offers a clearer narrative and potentially more actionable insights into the impact of ESGD. As explained in Section 3.3, further analyses are built on the initial findings and should examine *H2* from different perspectives. Their results, presented in the subsequent section, will also provide a basis to test the robustness of the main model. This progression will ensure that the interpretations in Section 5 are well-supported and offer a more comprehensive understanding of the impact of ESGD on ownership dynamics.

4.2.3. Regression Results: Further Analyses

The results of the further analysis provide several important insights into the relationship between ESGD, ESG performance, and institutional ownership. First, **lagged** variations of Model (V) examine institutional ownership levels over the five years after the observations of *ESGD* and *ESG* performance. Across all observed years, the signs of the relationships remain consistent: *ESGD* and *ESG* consistently show a positive association, and the coefficient for the interaction

Table 4: Univariate Regression Results

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
<i>Refinitiv ESGD</i>	0.068***	0.042**	0.164***	-0.003	0.072*	0.027	0.001
<i>Refinitiv ESG</i>		0.061***	0.230***	0.143***	0.242***	-0.006	-0.083*
<i>Refinitiv ESGD*ESG</i>			-0.003***		-0.002**	-0.0002	0.0004
<i>Log(TotAsts)</i>				-3.757***	-3.749***		0.440
<i>Beta</i>				-0.882	-0.909		0.044
<i>Lev</i>				0.048***	0.050***		-0.009
<i>EPS</i>				0.001	0.001		0.002
<i>Grwth</i>				-0.025**	-0.025*		0.009
<i>Log(TRVOL)</i>				1.408***	1.409***		
<i>ROA</i>				-0.035*	-0.034*		0.020
<i>MTB</i>				-0.009	-0.009		-0.006
<i>Fixed Effects</i>	Y, I, C	Y, F	Y, F				
<i>Constant</i>	25.719***	23.269***	15.820***	82.073***	77.278***		
<i>Observations</i>	5,229	5,229	5,229	4,733	4,733	5,229	4,770
<i>R²</i>	0.469	0.470	0.471	0.541	0.541	0.819	0.855
<i>Adjusted R²</i>	0.460	0.461	0.463	0.531	0.532	0.786	0.826
<i>Residual Std. Error</i>	15.466	15.444	15.427	14.041	14.036	9.735	8.557
<i>F Statistic</i>	53.983***	53.681***	53.323***	59.346***	58.795***		
<i>df</i>	5,144	5,143	5,142	4,640	4,639	4,420	3,978

*, **, *** Indicate statistical significance at the 5%, 1%, and 0.1% level, respectively

Coefficients for different univariate regressions on PctOwnInst as dependent variable: (I) includes only ESGD and fixed effects; (II) adds ESG performance; (III) adds the interaction term ESG*ESGD; (IV) adds controls, leaves out the interaction term ESG*ESGD; (V) adds the interaction term ESG*ESGD and controls; (VI) and (VII) add firm-fixed effects, with (VI) including the interaction term, and (VII) adding controls but excluding share volatility (treated as a quasi-fixed effect, as measured once per firm for the whole sample period; thus included in firm-fixed effects). Fixed effects abbreviations: Y = Year, I = Industry, C = Country of Headquarters, F = Firm. All variables are defined in Appendix 2.

Table 5: Comparison of Estimated Effects for Refinitiv and Bloomberg Measures

<i>Model</i>	<i>ESGD</i>		<i>ESG</i>		<i>ESGD * ESG</i>	
	<i>Refinitiv</i>	<i>Bloomberg</i>	<i>Refinitiv</i>	<i>Bloomberg</i>	<i>Refinitiv</i>	<i>Bloomberg</i>
(I)	+0.068***	+0.018				
(II)	+0.042**	+0.011	+0.061***	+0.436**		
(III)	+0.164***	+0.085**	+0.230***	+1.580***	-0.003***	-0.022**
(IV)	-0.003	-0.00003	+0.143***	+0.569***		
(V)	+0.072*	-0.015	+0.242***	+0.337	-0.002**	+0.004
(VI)	+0.027	+0.120**	-0.006	-0.047	-0.0002	-0.009
(VII)	+0.001	+0.060	-0.083*	-1.043	+0.0004	+0.003

*, **, *** Indicate statistical significance at the 5%, 1%, and 0.1% level, respectively

Summary of the coefficients for the main independent variables (ESGD, ESG, and ESGD*ESGD) across all seven univariate regressions of PctOwnInst, as presented in Section 3.3. Comparison of the results using Refinitiv versus Bloomberg measures. The detailed regression results are shown in Table 3 (Refinitiv) and Appendix 7 (Bloomberg). All variables are defined in Appendix 2.

term $ESGD * ESG$ is negative. In contrast, while ESG maintains its significance throughout the entire period, significance for $ESGD$ and $ESGD * ESG$ varies as it is only identified in the fourth year. The adjusted R-squared values range between 0.547 and 0.576 across the models.

The transformation of ESG , $ESGD$, and $PctOwnInst$ into Year-over-Year (YoY) **percentage changes** provides a different analytical perspective. However, regression models incorporating either same-year or lagged deltas¹⁰ reveal no significant associations: The estimators for $\Delta ESGD$, ΔESG , and their interaction term $\Delta ESGD * \Delta ESG$ predict positive but insignificant effects on institutional ownership changes. The adjusted R-squared value dropped to 0.002 and 0.004, respectively.

Splitting the data sample into groups with low and high stock trading volumes provides further differentiation. In firms with low stock turnover, negative coefficients are estimated for $ESGD$ (significant) and ESG (insignificant), while $ESGD * ESG$ is positive but insignificant. Conversely, in firms with high trading volumes, both $ESGD$ and ESG exhibit significant positive impacts on institutional ownership, whereas $ESGD * ESG$ becomes significantly negative. The adjusted R-squared values show a noticeable difference between the two groups, with 0.295 for the low liquidity group and 0.629 for the high liquidity group.

The **PVAR model** reveals a unidirectional relationship between $ESGD$ and institutional ownership. While institutional ownership is not predicted to significantly influence subsequent $ESGD$, $ESGD$ is assumed to exert a statistically significant positive effect on subsequent institutional ownership levels (coefficient of 0.0478, significant at the 5% level). In addition, $ESGD$ is also significantly connected to its own future levels (positive) and future ESG performance (negative). However, the findings suggest no significant reaction of $ESGD$ to earlier institutional ownership levels or ESG performance.

The **Difference-in-Difference** approach sheds light on the impact of substantial increases in $ESGD$ on institutional ownership. Companies in the treatment group, which encountered a notable rise in $ESGD$ during the sample period, experienced a smaller decline in institutional ownership compared to the control group with stable $ESGD$ levels: The average institutional ownership in the treatment group decreased by -5.0 percentage points (pp) from 24.5% in 2016 to 19.5% in 2022, while the control group faced a decline from 25.5% to 17.0% (-8.5pp).

Lastly, the analysis of **ownership concentration** shows no significant association between $ESGD$ and ownership structure. While the estimation predicts a positive but insignificant effect of $ESGD$, the coefficients of ESG and $ESGD * ESG$ have a negative insignificant sign. The adjusted R-squared for this model is 0.092.

These results underscore the complexity of the relationships between $ESGD$, ESG performance, and institutional ownership and emphasize the importance of considering various dimensions in the analysis. The models for further analyses are summarized in Appendix 8, with detailed results presented starting from Appendix 9.

5. Discussion and Implications

The upcoming section analyzes the results from Section 4 and explores their broader implications. Section 5.1 interprets the results, discussing how the findings relate to the initial hypotheses and existing literature. Section 5.2 evaluates the strengths and limitations of the study. It highlights the robustness of the findings while also acknowledging potential constraints. Section 5.3 explores the implications of the findings for various stakeholders, including companies, investors, and policymakers. It also suggests possible directions for future research in this area.

5.1. Interpretation of Findings

The findings from the multivariate regression in Section 4.2.1 largely support **hypothesis H1**, suggesting that $ESGD$ influences ownership structure in general. Across most model variations, the results were consistent, indicating a statistically **significant effect of $ESGD$** on the distribution of a firm's shares. Only Bloomberg's absolute $ESGD$ measure, in combination with country- and industry-fixed effects, appears to have a less significant impact on ownership structure. The overall trend aligns with the theoretical rationale. Prior research suggests that $ESGD$ serves two primary functions: On the one hand, it reduces information asymmetries, offering investors additional insights into the company (Dhaliwal et al., 2011, p. 90). On the other hand, it acts as a legitimizing tool, helping firms align with societal expectations and investor preferences (de Villiers & van Staden, 2006, pp. 763-764). The significant differences in ownership structures for varying $ESGD$ levels in Section 4.2.1 suggest that shareholders respond differently to the additional information and legitimacy $ESGD$ provides. This differential response among investors is consistent with the findings of Cho et al. (2013, p. 5), who emphasize that investors are not homogeneous in their degree of informedness. $ESGD$ contributes to bridging this informational gap, but not all investors process and value this information uniformly. Some investors, particularly those with greater resources and expertise, are better equipped to gather and utilize ESG data in their decision-making processes (Cho et al., 2013, p. 5). Even if different investor types would perceive the additional information identically, they might differ in how much importance they attribute to non-financial data. For instance, long-term investors may prioritize long-term sustainability practices (Qa'dan & Suwaidan, 2019, p. 35) and thus react more heavily by adjusting their shareholdings in reaction to new information about those practices. Consequently, the distinct characteristics among investor groups

¹⁰ The variation with one-year lagged deltas compares changes in $ESGD$ and ESG (between year $t-1$ and $t=0$) with ownership changes in the subsequent year (between $t=0$ and $t+1$). In contrast, the model with same-year deltas calculates all delta values between $t-1$ and $t=0$. The detailed models can be found in Appendix 7.

(see Section 2.2.1) lead to disparate responses to ESGD, which creates unique ownership structures and distributions of shareholdings. The more robust results observed for **relative** ESGD measures highlight that investors may be more concerned with how much a company discloses compared to its peers rather than the absolute volume of ESG information disclosed. This could suggest that ESGD is particularly influential during investment decision-making, where investors select between competing firms. Firms that disclose more ESG information relative to their industry peers may be viewed as more transparent, which fosters investors' confidence and, thus, stock liquidity (Healy & Palepu, 2001, p. 429). This greater liquidity leads to higher stock- and investor turnover, which potentially shifts the composition of the investor base.

However, it is essential to acknowledge that the multivariate regression approach does not allow for a deeper quantitative exploration of the association. While it indicates whether ESGD has an impact in general, it does not provide insight into the direction and the actual strength, **limiting the interpretability** of the results. As a direct consequence, it remains unclear to what extent the significant interaction term might offset the potential effect of ESGD. While significance is observed, the specific direction or magnitude remains ambiguous. Moreover, the model offers no information about the impact of ESGD on the particular ownership types, leaving unanswered questions about how ESGD might influence the distribution of shares to explicit investor groups. As a result, it is not possible to connect these findings to the literature on specific ownership types and address the theory and empirical evidence presented in Section 2.4. Nonetheless, these limitations do not oppose the primary goal of *H1*, as this hypothesis was intended to explore the broader relevance of ESGD for ownership structure. This investigation lays the groundwork for a more detailed analysis of institutional ownership and its specific relationship with ESGD.

Following this foundation, the primary analysis in Section 4.2.2 examines the impact of ESGD on institutional ownership and reveals mostly consistent support for **hypothesis H2**: For the **relative ESGD measure**, the univariate regression models indicate a **significant positive** effect of ESGD on institutional ownership levels in most model versions. This result suggests that institutional investors, known to play a pivotal role in corporate governance and monitoring (Serafeim, 2015, p. 36), are responsive to ESGD. The positive impact of ESGD appears to be particularly strong for firms that are typically more attractive to short-term institutional investors: Applied to a subgroup of companies with high trading volumes, the regression model shows an even stronger association between ESGD and institutional ownership than on the entire data sample.

The general positive effect aligns with the expectations of the academic literature. As discussed in Section 2.4, there are multiple explanations for this relationship. Dhaliwal et al. (2011, p. 80) argue that ESGD attracts long-term institutional investors by facilitating their monitoring and governance role. Another rationale might be the reduced infor-

mation asymmetries among investors because of increased transparency, resulting in higher stock liquidity. This liquidity is a critical factor for institutional investors with short-term horizons, who prefer to invest in markets where they can trade easily and efficiently (Diamond and Verrecchia, 1991, pp. 1326-1327; Healy and Palepu, 2001, p. 429). Legitimacy theory provides another perspective: by disclosing ESG performance, companies demonstrate alignment with societal values, which builds stakeholder trust and appeals to investors who prioritize social responsibility. This transparency signals a commitment to sustainable practices, making the company more attractive to socially conscious institutional investors (de Villiers and van Staden, 2006, p. 764; Suchman, 1995, p. 574). Even though the analysis gives no insights into the causes of this relationship, those arguments could explain the behavior of institutional investors.

While the results from Section 4.2.2 indicate a positive impact of ESGD on institutional ownership, they also suggest this influence may not stem from entirely new information for institutional investors: The **negative interaction term** between ESGD and ESG performance indicates that the effect of ESGD weakens as firms' ESG performance improves. This finding contradicts the result from Dhaliwal et al. (2011, p. 80) that higher transparency has a stronger impact on institutional ownership at higher levels of ESG performance. Instead, it may reflect that institutional investors, already possessing a high level of information about the firms' ESG profiles, are less surprised by the content of the disclosed information. This is likely due to their naturally superior information level (Cho et al., 2013, p. 5). Moreover, the negative interaction term suggests that institutional investors might view the disclosure act itself, rather than its specific ESG content, as valuable. An alternative explanation for the negative interaction might be that institutional investors exploit reduced information asymmetries in other investor groups to their advantage: As more information becomes available to other, less informed market participants, the demand for shares of the good-performing firms increases. This higher demand is driven by the growing preference for companies with strong ESG performance (Morgan Stanley Institute for Sustainable Investing, 2024) and higher stock liquidity, which results from greater transparency (Healy & Palepu, 2001, p. 429). According to the principles of supply and demand, the higher interest in good-performing companies would enable institutional investors to sell the respective shares at higher prices (Marshall, 2013, p. 290 et seq.). This strategic behavior would explain the weaker association between ESGD and institutional ownership in firms with better ESG performance. Generally, the negative interaction term could imply that firms with strong ESG performance may not need to disclose as aggressively to attract institutional investors. For these firms, a targeted, strategic disclosure approach focusing on aspects less visible or quantifiable to external investors may be more effective. In contrast, firms with weaker ESG performance might use additional disclosures to legitimize their practices and address concerns about sustainability. Therefore, these firms might

still attract institutional investors, even if their ESG performance remains relatively low.

Overall, the robustness checks (see Appendix 15) indicate a strong and consistent trend, further validating the findings on the impact of ESGD and its interaction term with ESG performance. However, the previous findings are inconsistent when using the **absolute ESGD** measures from Bloomberg, where the coefficients are weaker and less robust. This difference may indicate that institutional investors take a relative view of ESGD, evaluating a firm's disclosure performance by comparing it to similar firms in its industry rather than focusing solely on absolute ESGD levels. By placing ESGD in context, investors may prioritize investment opportunities that stand out positively within their sector rather than those meeting an absolute standard across industries. This approach allows institutional investors to assess a firm's ESG transparency compared to its peers, making the evaluation more relevant to the company's competitive position. Alternatively, the inconsistent results may indicate that the effect of ESGD on institutional ownership is less reliable than initially suggested.

Questions about the **reliability** are also raised by the mixed results of the additional analyses in Section 4.2.3. The difference-in-difference approach supports a positive impact of ESGD on institutional ownership by showing that ESGD diminishes the decline in institutional ownership between 2016 and 2022. In contrast, the results for lagged ownership levels and percentage changes in ownership were primarily positive but not significant. This insignificance suggests that ESGD does not lead to an immediate or measurable reaction from institutional investors in the form of portfolio adjustments, challenging the assumption that ESGD drives active changes in ownership structures. Instead, ESG transparency tends to influence only the long-term orientation of institutional investors rather than triggering immediate buy or sell decisions. This lack of visible short-term reallocation weakens the overall suggested impact of ESGD on institutional ownership, indicating that the increased transparency may not be as influential as some literature suggests. These findings are supported by the analysis of ownership concentration, which is often linked to institutional investors as they are associated with higher total ownership stakes and concentrated ownership (Diamond and Verrecchia, 1991, p. 1327; Hoq et al., 2010, p. 24). The results show an insignificant positive coefficient for ESGD and encourage that the immediate impact of ESG transparency on ownership structure is limited.

The mixed findings also raise questions about the **bi-directional relationship** between ESGD and institutional ownership. Ntim and Soobaroyen (2013, p. 472) suggest that institutional shareholders support higher disclosure to project a socially responsible image. Furthermore, they might push for ESGD to reduce information asymmetries between the company and other shareholders (Ntim & Soobaroyen, 2013, p. 472). Therefore, the direction of causality remains ambiguous, with theoretical explanations supporting both possibilities (Healy & Palepu, 2001, p. 431). A

PVAR analysis was conducted to address this, and possible relationships in both directions were examined. The results indicate an effect of ESGD on ownership but no significant impact of ownership on ESGD, which contradicts literature such as Ramdhony et al. (2024, p. 531). The discrepancy with Ramdhony et al. (2024, p. 531) might stem from the different investor types analyzed, as their study focused on government-, director-, and block ownership rather than institutional investors. Additionally, their model was applied to a sample of Mauritian firms in an emerging economy context, whereas this study examines companies from primarily developed countries.

In summary, while evidence supports the positive impact of ESGD on institutional ownership, the findings are complex and not fully robust across different perspectives and measurements.

5.2. Strengths and Limitations

In contrast to the ambiguity in the further analyses of Section 4.2.3, the main model in Section 4.2.2 demonstrates strong **robustness** and meets various **quality criteria** (see Appendix 15). Firstly, several models and methods support the main model's validity and findings: The model without outliers, Lasso Regression, and models with Box-Cox and log transformations show mostly consistent signs and significance of the coefficients compared to Model (V). For the reliability of the main regression, a VIF test for multicollinearity, the Breusch-Pagan test for homoscedasticity, and the Durbin-Watson test for autocorrelation provide valid results. Furthermore, techniques such as bootstrapping and coefficient tests using robust standard errors enhance the reliability of the results and minimize uncertainty: The confidence interval is predominantly positive, and the robust coefficient for *ESGD* is identical to the estimate in Model (V). Additionally, objectivity is supported by automated techniques like Lasso regression and bootstrapping, which help to eliminate subjective model assumptions that could be introduced by the author. The main model's performance is further confirmed by a high R-squared value, particularly in comparison to similar studies, and indicates a strong explanatory power. The analysis benefits from a large sample size that spans an extensive, more recent period, primarily within the context of a developed economy. This more recent time frame is particularly relevant to the current dynamics of ESG transparency, enabling a more comprehensive investigation than existing studies. Multiple models account for different perspectives, striving to develop a well-rounded understanding of the relationship between ESGD and institutional ownership. However, this comprehensive picture is not entirely achieved because of the inconsistency in the results of further analyses, representing a significant **limitation** for the overall meaningfulness of the findings. In addition, even with an extended time frame compared to existing literature, the study may still fall short in capturing the reactions of investors with long-term horizons, such as institutional investors. The sample period of six years, from 2016 to 2022, might not be long enough to fully observe delayed or gradual responses.

Because of those longer investment horizons, investors may not adjust their portfolios quickly in response to increased ESGD but react slower than the six years covered in this study. Moreover, while the univariate regression examines the effect of ESGD on institutional ownership in isolation, its design ignores the interrelated nature of ownership dynamics. Any change in institutional ownership is inherently linked to shifts in the shares held by other investor types, which may lead to positive or negative consequences depending on their unique characteristics. Thus, while insights can be drawn regarding institutional ownership, it is crucial to consider the simultaneous changes occurring among other shareholder groups. This interrelation also includes the treatment of the free float, particularly shares held by private or retail investors who are not captured in the Orbis database and, therefore, not included in any analysis. This exclusion is justified, as including estimations of these shares could introduce distortions due to the potential double counting of shareholding links within Orbis, making the shareholdings not additive to 100%. Double counting generally poses a significant weakness in the ownership data, potentially leading to inaccurate results. In contrast to the potential ambiguities regarding the design of ownership variables, the construction of the ESGD measures used in this study is precise and incorporates different perspectives (absolute vs. relative). However, a notable gap in those measures is the failure to differentiate between mandatory and voluntary disclosures. This distinction could offer further insights for companies, particularly in determining whether to enhance their disclosures beyond the levels required by existing regulatory frameworks.

5.3. Implications and Outlook

The evidence from this study suggests that ESGD impacts a company's ownership structure, though it is limited to relative ESGD and reflects some inconsistencies in the results. The findings indicate that companies can enhance their appeal to institutional investors through increased ESGD. This idea is particularly relevant for firms with historically poor ESG performance that have been reluctant to share such information. By increasing transparency, these companies may attract institutional investors even though the disclosed information reveals negative facts about the company. While the results in this paper suggest a positive connection to institutional ownership, it is essential to recognize that the relationship is not universally applicable. Future research could benefit from differentiating between various types of institutional investors, such as hedge funds, pension funds, and banks, as their investment objectives and strategies may differ significantly. Additionally, considering the varying time horizons among these investors could provide more nuanced insights into how ESGD influences institutional ownership. Moreover, as highlighted by Paul Polman, the former CEO of Unilever, companies should aim to attract a shareholder base that aligns with their strategic goals. This perspective underlines that institutional investors should not be regarded as

the ultimate solution for a company's success. Thus, expanding the focus beyond institutional investors to examine how ESGD influences other types of investors or investor characteristics might help companies shape a shareholder base that is more aligned with their strategy. One promising avenue would be to examine the distinctions between mandatory and voluntary disclosures, particularly regarding regulatory frameworks and how they shape corporate behavior. In environments with strict mandatory disclosure requirements, companies have less flexibility to adapt their practices strategically. In contrast, voluntary disclosure settings offer more freedom, allowing companies to shape their transparency and influence ownership dynamics to attract various types of investors (Lang & Lundholm, 1996, p. 468). Those findings could provide more practical guidance for companies and insights for regulators about the different effects of mandatory and voluntary disclosure of ESG information. Additionally, extending the analysis to encompass a longer time horizon may help capture the long-term reactions of investors to ESGD. Furthermore, future studies could explore possible connected effects of ESGD and institutional ownership on a company's financials. As the theory assumes a positive economic impact for increased disclosure and institutional ownership (see Section 2.1.2 and 2.2.2), a potential interaction effect between both factors might further lever those financial benefits.

A similar approach is already taken by Ntim and Soobaroyen (2013, p. 484), who investigate the moderating effect of CSR disclosure on corporate performance through ownership variables. They suggest a negative but insignificant impact of institutional ownership on corporate performance for increasing CSR levels. Expanding this approach could provide insights into how companies can capitalize on the indirect effect of increased institutional ownership through higher ESG transparency.

6. Conclusion

This paper investigates if ESGD impacts a company's ownership structure. Essential insights are drawn from a comprehensive examination of the relationship between ESGD and ownership dynamics, contributing to the growing body of literature on sustainable investing. The findings indicate a significant effect of ESGD on ownership structure in general, suggesting that the nature and extent of ESG transparency influence ownership dynamics. Further quantitative interpretation is limited due to the multivariate research design employed. However, the results suggest that the general impact of ESGD on ownership structure is meaningful and warrants deeper investigation in future research endeavors. One of the most critical areas identified and examined in this study is the role of institutional investors and their reaction to ESGD. The analysis conducted in this paper reveals that, while not entirely consistent, the estimated effect of ESGD on institutional ownership is significantly positive, accompanied by a significant negative interaction term with ESG performance. This

finding suggests that institutional investors place considerable importance on ESG transparency, even when the actual content of those disclosures may not reflect high ESG performance. Theory often explains such trends with the monitoring responsibilities of institutional investors, who require comprehensive information to assess potential risks. Interestingly, the results imply that additional disclosure of negative ESG performance is not less attractive to institutional investors than information about positive ESG performance. Given their existing information advantage, institutional investors may already possess insights into ESG performance that diminish the surprising effect of damaging disclosures. Furthermore, the analysis highlights differences between absolute and relative measures of ESGD. Institutional investors appear to prioritize relative ESG transparency – evaluating a company’s disclosures in relation to its peers – over absolute disclosure levels. This finding indicates that relative comparisons of disclosure levels among similar companies provide a more consistent and precise understanding of how ESG transparency is perceived in the context of industry standards. Despite these valuable insights, the study has its limitations. The results reveal some inconsistencies or contradictions, indicating a need for further investigation. Additionally, the time horizon of the study, spanning a maximum of six years, may not be long enough to capture investors’ comprehensive reaction to changes in ESG disclosures. Potential distortions arising from the variable construction, such as double counting of ownership links and the lack of differentiation between mandatory and voluntary ESGD, further complicate the analysis.

To build on the findings of this research, future studies should focus on several key areas. First, a deeper examination of the distinctions between voluntary and mandatory disclosures, particularly in the context of regulatory frameworks, could provide valuable insights into how different disclosure types affect ownership structures. Furthermore, future research could differentiate institutional ownership and examine the impact of ESGD on specific types of institutional investors, such as hedge funds, pension funds, and banks, while also considering their varying investment horizons. Additionally, exploring other ownership types, such as individual or government investors, and concentration levels may yield a more nuanced understanding of the interaction between ESG transparency and ownership dynamics. Finally, investigating the interaction effect between ESGD and institutional ownership on company performance would be beneficial, adding further depth to this critical area of research.

In conclusion, this thesis offers a foundational exploration of a contemporary topic with the potential for valuable insights for companies, legislators, and investors. The interplay between ESGD and ownership structure is increasingly relevant as the demand for transparency and accountability in corporate governance rises. This study opens avenues for further research that can enhance understanding and inform practices in sustainable investing.

References

- Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, 17(2), 223–249. [https://doi.org/10.1016/0304-405X\(86\)90065-6](https://doi.org/10.1016/0304-405X(86)90065-6)
- Barnea, A., & Rubin, A. (2010). Corporate Social Responsibility as a Conflict Between Shareholders. *Journal of Business Ethics*, 97(1), 71–86. <https://doi.org/10.1007/s10551-010-0496-z>
- Belkhir, M., Saad, M., & Samet, A. (2020). Stock extreme illiquidity and the cost of capital. *Journal of Banking & Finance*, 112, 105281. <https://doi.org/10.1016/j.jbankfin.2018.01.005>
- Berk, J., & DeMarzo, P. (2023). *Corporate Finance, Global Edition*. Pearson Education, Limited. <http://ebookcentral.proquest.com/lib/munchentech/detail.action?docID=7261465>
- Bernow, S., Godsall, J., Klemmner, B., & Merten, C. (2019). More than values: The value-based sustainability reporting that investors want. Retrieved November 5, 2024, from <https://www.mckinsey.com/capabilities/sustainability/our-insights/more-than-values-the-value-based-sustainability-reporting-that-investors-want#/>
- Beyer, A., Larcker, D. E., & Tayan, B. (2014). *Study on How Investment Horizon and Expectations of Shareholder Base Impact Corporate Decision-Making* (Report). National Investor Relations Institute and Stanford University. <https://www.gsb.stanford.edu/faculty-research/publications/2014-study-how-investment-horizon-expectations-shareholder-base-impact>
- Bishara, M. K., Andrikopoulos, P., & Eldomiatiy, T. (2020). Ownership structure, information asymmetry and growth of the firm: Implications from nonfinancial firms listed in S&P500. *Managerial & Decision Economics*, 41(8), 1580–1589. <https://doi.org/10.1002/mde.3204>
- Bloomberg. (2024). Bloomberg Professional. Retrieved November 7, 2024, from <https://www.bloomberg.com/professional/>
- Brammer, S., & Pavelin, S. (2008). Factors influencing the quality of corporate environmental disclosure. *Business Strategy and the Environment*, 17(2), 120–136. <https://doi.org/10.1002/bse.506>
- Brush, T. H., Bromiley, P., & Hendrickx, M. (2000). The free cash flow hypothesis for sales growth and firm performance. *Strategic management journal*, 21(4), 455–472.
- Bureau van Dijk. (2018). *Orbis data guide* (Report).
- Bushee, B. J. (1998). The Influence of Institutional Investors on Myopic R&D Investment Behavior. *The Accounting Review*, 73(3), 305–333. <http://www.jstor.org/stable/248542>
- Bushee, B. J., & Noe, C. F. (2000). Corporate Disclosure Practices, Institutional Investors, and Stock Return Volatility. *Journal of Accounting Research*, 38, 171–202. <https://doi.org/10.2307/2672914>
- Bushman, R. M., Piotroski, J. D., & Smith, A. J. (2004). What Determines Corporate Transparency? *Journal of Accounting Research (Wiley-Blackwell)*, 42(2), 207–252. <https://doi.org/10.1111/j.1475-679X.2004.00136.x>
- Cahan, S. F., De Villiers, C., Jeter, D. C., Naiker, V., & Van Staden, C. J. (2016). Are CSR Disclosures Value Relevant? Cross-Country Evidence. *European Accounting Review*, 25(3), 579–611. <https://doi.org/10.1080/09638180.2015.1064009>
- Cao, F., Peng, S. S., & Ye, K. (2019). Multiple Large Shareholders and Corporate Social Responsibility Reporting. *Emerging Markets Review*, 38, 287–309. <https://doi.org/10.1016/j.ememar.2019.02.004>
- Caselli, S., Gatti, S., Chiarella, C., Gigante, G., & Negri, G. (2023). Do shareholders really matter for firm performance? Evidence from the ownership characteristics of Italian listed companies. *International Review of Financial Analysis*, 86, 102544. <https://doi.org/10.1016/j.irfa.2023.102544>
- Cheng, B., Ioannou, I., & Serafeim, G. (2014). Corporate social responsibility and access to finance. *Strategic Management Journal*, 35(1), 1–23. <https://doi.org/10.1002/smj.2131>
- Cho, S. Y., Lee, C., & Pfeiffer, R. J. (2013). Corporate social responsibility performance and information asymmetry. *Journal of Accounting & Public Policy*, 32(1), 71–83. <https://doi.org/10.1016/j.jaccpu.2012.10.005>
- Christensen, D. M., Serafeim, G., & Sikochi, A. (2022). Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings. *Accounting Review*, 97(1), 147–175. <https://doi.org/10.2308/TAR-2019-0506>

- Christensen, H. B., Hail, L., & Leuz, C. (2021). Mandatory CSR and sustainability reporting: economic analysis and literature review. *Review of Accounting Studies*, 26(3), 1176–1248. <https://doi.org/10.1007/s11142-021-09609-5>
- Ciaburro, G. (2018). Regression Analysis with R : Design and develop statistical nodes to identify unique relationships within data at scale. <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1703784>
- Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2008). Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. *Accounting, Organizations and Society*, 33(4), 303–327. <https://doi.org/10.1016/j.aos.2007.05.003>
- Cleveland, S., Sullivan, K., Poole, V., & Chahed, Y. (2023). Overcoming the hurdles to integrating sustainability into business strategy. *Deloitte Insights Magazine*, 32.
- Code of Federal Regulations. (2023). Title 17, Part 210: Form and content of and requirements for financial statements, securities acts amendments. Retrieved November 4, 2024, from <https://www.ecfr.gov/current/title-17/chapter-II/part-210>
- Cornett, M. M., Marcus, A. J., Saunders, A., & Tehranian, H. (2007). The impact of institutional ownership on corporate operating performance. *Journal of Banking & Finance*, 31(6), 1771–1794. <https://doi.org/10.1016/j.jbankfin.2006.08.006>
- DataTracks. (2024). A Complete Guide to ESG Reporting in the US. Retrieved November 4, 2024, from <https://www.datatracks.com/us/blog/esg-reporting-in-the-us/>
- Dattalo, P. (2013). Analysis of Multiple Dependent Variables. <https://doi.org/10.1093/acprof:oso/9780199773596.001.0001>
- de Villiers, C., & van Staden, C. J. (2006). Can less environmental disclosure have a legitimising effect? Evidence from Africa. *Accounting, Organizations & Society*, 31(8), 763–781. <https://doi.org/10.1016/j.aos.2006.03.001>
- Deegan, C. (2002). The legitimising effect of social and environmental disclosures—A theoretical foundation. *Accounting, Auditing & Accountability Journal*, 15, 282–311. <https://doi.org/10.1108/0951357021043585>
- Demsetz, H. (1983). The structure of ownership and the theory of the firm. *The Journal of law and economics*, 26(2), 375–390.
- Dhaliwal, D. S., Heitzman, S., & Zhen Li, O. (2006). Taxes, Leverage, and the Cost of Equity Capital. *Journal of Accounting Research*, 44(4), 691–723. <https://doi.org/10.1111/j.1475-679X.2006.00214.x>
- Dhaliwal, D. S., Oliver Zhen, L., Tsang, A., & Yong George, Y. (2011). Voluntary Nonfinancial Disclosure and the Cost of Equity Capital: The Initiation of Corporate Social Responsibility Reporting. *Accounting Review*, 86(1), 59–100. <https://doi.org/10.2308/accr.00000005>
- Diamond, D. W., & Verrecchia, R. E. (1991). Disclosure, Liquidity, and the Cost of Capital. *Journal of Finance (Wiley-Blackwell)*, 46(4), 1325–1359. <https://doi.org/10.1111/j.1540-6261.1991.tb04620.x>
- Dowling, J., & Pfeffer, J. (1975). Organizational legitimacy: Social values and organizational behavior. *Pacific sociological review*, 18(1), 122–136.
- Draper, N. (1998). *Applied regression analysis*. McGraw-Hill. Inc.
- Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management science*, 60(11), 2835–2857.
- Eccles, R. G., Serafeim, G., & Krzus, M. P. (2011). Market Interest in Nonfinancial Information. *Journal of Applied Corporate Finance*, 23(4), 113–127. <https://doi.org/10.1111/j.1745-6622.2011.00357.x>
- Eng, L. L., & Mak, Y. T. (2003). Corporate governance and voluntary disclosure. *Journal of Accounting and Public Policy*, 22(4), 325–345. [https://doi.org/10.1016/S0278-4254\(03\)00037-1](https://doi.org/10.1016/S0278-4254(03)00037-1)
- European Commission. (2024). Corporate sustainability reporting. Retrieved November 18, 2024, from https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en
- European Parliament and Council of the European Union. (2002). Regulation (EC) No 1606/2002. Retrieved November 4, 2024, from <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32002R1606>
- European Parliament and Council of the European Union. (2014). Directive 2014/95/EU. Retrieved November 4, 2024, from <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=celex%3A32014L0095>
- European Parliament and Council of the European Union. (2022). Directive (EU) 2022/2464. Retrieved November 4, 2024, from <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022L2464>
- Friedman, H., & Heinle, M. (2016). Taste, information, and asset prices: implications for the valuation of CSR. *Review of Accounting Studies*, 21(3), 740–767. <https://doi.org/10.1007/s11142-016-9359-x>
- Gebhardt, W. R., Lee, C. M. C., & Swaminathan, B. (2001). Toward an Implied Cost of Capital. *Journal of Accounting Research*, 39(1), 135–176. <https://doi.org/10.1111/1475-679X.00007>
- German Commercial Code. (2024). Subdivision 4 - Disclosure. Verification by the operator of the Federal Gazette. Retrieved November 4, 2024, from https://www.gesetze-im-internet.de/englisch_hgb/englisch_hgb.html#p0872
- Global Reporting Initiative. (2024). GRI Standards. Retrieved November 5, 2024, from <https://www.globalreporting.org/standards/>
- Governance & Accountability Institute. (2024). Sustainability Reporting in Focus. Retrieved November 4, 2024, from <https://www.ga-institute.com/research/ga-research-directory/sustainability-reporting-trends/2024-sustainability-reporting-in-focus/>
- Grewal, J., Riedl, E. J., & Serafeim, G. (2019). Market Reaction to Mandatory Nonfinancial Disclosure. *Management Science*, 65(7), 3061–3084. <https://doi.org/10.1287/mnsc.2018.3099>
- Haniffa, R. M., & Cooke, T. E. (2002). Culture, Corporate Governance and Disclosure in Malaysian Corporations. *Abacus*, 38(3), 317–349. <https://doi.org/10.1111/1467-6281.00112>
- Härdle, W. K., Simar, L., & Fengler, M. R. (2024). *Applied Multivariate Statistical Analysis* (Sixth Edition). Springer. <https://doi.org/10.1007/978-3-031-63833-6>
- Hasseldine, J., Salama, A. I., & Toms, J. S. (2005). Quantity versus quality: the impact of environmental disclosures on the reputations of UK Plcs. *The British Accounting Review*, 37(2), 231–248. <https://doi.org/10.1016/j.bar.2004.10.003>
- Healy, P. M., Hutton, A. P., & Palepu, K. G. (1999). Stock Performance and Intermediation Changes Surrounding Sustained Increases in Disclosure. *Contemporary Accounting Research*, 16(3), 485–520. <https://doi.org/10.1111/j.1911-3846.1999.tb00592.x>
- Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(1), 405–440. [https://doi.org/10.1016/S0165-4101\(01\)00018-0](https://doi.org/10.1016/S0165-4101(01)00018-0)
- Hirst, S., Kastiel, K., & Kricheli-Katz, T. (2023). How Much Do Investors Care About Social Responsibility? *Wisconsin Law Review*, 2023(3), 977–1039. <https://doi.org/10.59015/wlr.DIHK9019>
- Hoffmann, C., Schiereck, D., & Zerfaß, A. (2022). *Handbuch investor relations und Finanzkommunikation*. Springer.
- Hoq, M. Z., Saleh, M., Zubayer, M., & Mahmud, K. T. (2010). The Effect of CSR Disclosure on Institutional Ownership. *Pakistan Journal of Commerce & Social Sciences*, 4(1), 22–39. <https://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=49791755&site=ehost-live>
- IFRS Foundation. (2024). Use of IFRS standards by jurisdiction. Retrieved November 4, 2024, from <https://www.ifrs.org/use-around-the-world/use-of-ifrs-standards-by-jurisdiction/#use-of-ifrs-accounting-standards-by-jurisdiction>
- International Organization for Standardization. (2024). What we do. Retrieved November 4, 2024, from <https://www.iso.org/what-we-do.html>
- ISIN Organization. (2024). ISIN. Retrieved November 14, 2024, from <https://www.isin.org/isin/>
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305–360.
- Jia, C., Ding, S., Li, Y., & Wu, Z. (2009). Fraud, Enforcement Action, and the Role of Corporate Governance: Evidence from China. *Journal of*

- Business Ethics*, 90(4), 561–576. <https://doi.org/10.1007/s10551-009-0061-9>
- Jiang, H., & Habib, A. (2009). The impact of different types of ownership concentration on annual report voluntary disclosures in New Zealand. *Accounting Research Journal*, 22(3), 275–304. <https://doi.org/10.1108/10309610911005590>
- Jouida, S. (2019). Bank capital structure, capital requirements and SRISK across bank ownership types and financial crisis: panel VAR approach. *Review of Quantitative Finance & Accounting*, 53(1), 295–325. <https://doi.org/10.1007/s11156-018-0750-5>
- Kalay, A. (2015). Investor sophistication and disclosure clienteles. *Review of Accounting Studies*, 20(2), 976–1011. <https://doi.org/10.1007/s11142-015-9317-z>
- Khlif, H., Ahmed, K., & Souissi, M. (2017). Ownership structure and voluntary disclosure: A synthesis of empirical studies. *Australian Journal of Management (Sage Publications Ltd.)*, 42(3), 376–403. <https://doi.org/10.1177/0312896216641475>
- Kim, J.-B., Li, B., & Liu, Z. (2018). Does social performance influence breadth of ownership? *Journal of Business Finance & Accounting*, 45(9/10), 1164–1194. <https://doi.org/10.1111/jbfa.12341>
- Koller, T. (2020). Warren Buffett: An appreciation. Retrieved November 13, 2024, from <https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/warren-buffett-an-appreciation>
- Lang, M. H., & Lundholm, R. J. (1996). Corporate Disclosure Policy and Analyst Behavior. *The Accounting Review*, 71(4), 467–492. <http://www.jstor.org/stable/248567>
- London Stock Exchange Group. (2024). Data and analytics. Retrieved November 7, 2024, from <https://www.lseg.com/en/data-analytics>
- Marshall, A. (2013). *Principles of economics*. Springer.
- Mathews, M. (1993). *Socially Responsible Accounting*. Chapman & Hall.
- Milne, M. J., & Patten, D. M. (2002). Securing organizational legitimacy. *Accounting, Auditing & Accountability Journal*, 15(3), 372–405. <https://doi.org/10.1108/09513570210435889>
- Moody's Analytics. (2024). Orbis. Retrieved November 7, 2024, from <https://www.moody.com/web/en/us/capabilities/company-reference-data/orbis.html>
- Morgan Stanley Institute for Sustainable Investing. (2024). Sustainable Signals: Understanding Individual Investors' Interests and Priorities. Retrieved November 4, 2024, from <https://www.morganstanley.com/ideas/sustainable-investing-on-the-rise>
- Moss, A., Naughton, J. P., & Wang, C. (2024). The Irrelevance of Environmental, Social, and Governance Disclosure to Retail Investors. *Management Science*, 70(4), 2626–2644. <https://doi.org/10.1287/mnsc.2023.4822>
- Ntim, C. G., & Soobaroyen, T. (2013). Corporate Governance and Performance in Socially Responsible Corporations: New Empirical Insights from a Neo-Institutional Framework. *Corporate Governance: An International Review*, 21(5), 468–494. <https://doi.org/10.1111/corg.12026>
- O'Dwyer, L. M., & Bernauer, J. A. (2014). *Quantitative Research for the Qualitative Researcher*. SAGE Publications, Inc.
- Patten, D. M. (2002). The relation between environmental performance and environmental disclosure: a research note. *Accounting, Organizations and Society*, 27(8), 763–773. [https://doi.org/10.1016/S0361-3682\(02\)00028-4](https://doi.org/10.1016/S0361-3682(02)00028-4)
- Plümper, T., & Troeger, V. E. (2019). Not so harmless after all: The fixed-effects model. *Political Analysis*, 27(1), 21–45.
- Polman, P. (2012). Unilever's CEO on Making Responsible Business Work. Retrieved October 30, 2024, from <https://hbr.org/podcast/2012/05/unilevers-ceo-on-making-respon>
- PwC. (2023). PwC's Global Investor Survey 2023. Retrieved November 4, 2024, from <https://www.pwc.com/gx/en/issues/c-suite-insights/global-investor-survey.html>
- Qa'dan, M. B. A., & Suwaidan, M. S. (2019). Board composition, ownership structure and corporate social responsibility disclosure: the case of Jordan. *Social Responsibility Journal*, 15(1), 28–46.
- Ramdhony, D., Gunensee, S., Mooneeapen, O., & Boolaky, P. (2024). CSR disclosure and ownership structure: insights from a dynamic empirical framework using an emerging economy context. *Journal of Applied Accounting Research*, 25(3), 524–546. <https://doi.org/10.1108/JAAR-12-2021-0338>
- Roberts, R. W. (1992). Determinants of corporate social responsibility disclosure: An application of stakeholder theory. *Accounting, Organizations and Society*, 17(6), 595–612. [https://doi.org/10.1016/0361-3682\(92\)90015-K](https://doi.org/10.1016/0361-3682(92)90015-K)
- Said, R., Zainuddin, Y. H., & Haron, H. (2009). The relationship between corporate social responsibility disclosure and corporate governance characteristics in Malaysian public listed companies. *Social Responsibility Journal*, 5(2), 212–226. <https://doi.org/10.1108/1747110910964496>
- Serafeim, G. (2015). Integrated Reporting and Investor Clientele. *Journal of Applied Corporate Finance*, 27(2), 34–51. <https://doi.org/10.1111/jacf.12116>
- Shleifer, A., & Vishny, R. W. (1997). A Survey of Corporate Governance. *Journal of Finance*, 52(2), 737–783. <https://doi.org/10.1111/j.1540-6261.1997.tb04820.x>
- Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Management Review*, 20(3), 571–610. <https://doi.org/10.5465/AMR.1995.9508080331>
- Sustainability Accounting Standards Board. (2024). SASB standards. Retrieved November 5, 2024, from <https://sasb.ifrs.org/standards/>
- Tagesson, T., Blank, V., Broberg, P., & Collin, S.-O. (2009). What explains the extent and content of social and environmental disclosures on corporate websites: a study of social and environmental reporting in Swedish listed corporations. *Corporate Social Responsibility and Environmental Management*, 16(6), 352–364. <https://doi.org/10.1002/csr.194>
- Thomsen, S., & Conyon, M. (2012). *Corporate Governance: Mechanisms and Systems*. McGraw-Hill.
- Thomsen, S., & Pedersen, T. (2000). Ownership structure and economic performance in the largest European companies. *Strategic Management Journal*, 21(6), 689. [https://doi.org/10.1002/\(SICI\)1097-0266\(200006\)21:6<689::AID-SMJ115>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-0266(200006)21:6<689::AID-SMJ115>3.0.CO;2-Y)
- Transparency International. (2023). Corruption Perceptions Index 2023. Retrieved November 7, 2024, from <https://www.transparency.org/en/cpi/2023>
- U.S. Securities and Exchange Commission. (2024). Regulation S-X, Article 14: The Enhancement and Standardization of Climate-Related Disclosures for Investors. Retrieved November 4, 2024, from <https://www.sec.gov/files/rules/final/2024/33-11275.pdf>
- Verrecchia, R. E. (1983). Discretionary disclosure. *Journal of Accounting and Economics*, 5, 179–194. [https://doi.org/10.1016/0165-4101\(83\)90011-3](https://doi.org/10.1016/0165-4101(83)90011-3)
- Wahl, M. (2006). The Ownership Structure of Corporations: Owners Classification & Typology. *EBS Review*, (21), 94–103. <https://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=25554803&site=ehost-live>
- Woodward, D., Edwards, P., & Birkin, F. (2001). Some Evidence on Executives' views of Corporate Social Responsibility. *The British Accounting Review*, 33(3), 357–397. <https://doi.org/10.1006/bare.2001.0165>
- Wooldridge, J. M. (2020). *Introductory Econometrics* (Seventh edition). Cengage Learning.
- Yu, K., Du, S., & Bhattacharya, C. (2013). Everybody's talking but is anybody listening? Stock market reactions to corporate social responsibility communications. *The Sustainability and The Corporation: Big Ideas*, 1–50.
- Zairis, G., Liargovas, P., & Apostolopoulos, N. (2024). Sustainable Finance and ESG Importance: A Systematic Literature Review and Research Agenda. <https://doi.org/10.3390/su16072878>
- Zhong, M., & Gao, L. (2017). Does corporate social responsibility disclosure improve firm investment efficiency? *Review of Accounting & Finance*, 16(3), 348–365. <https://doi.org/10.1108/RAF-06-2016-0095>



ESG Regulation Across the Globe: Does ESG Regulation Pay Off?

Antonia Engel

Technical University of Munich

Abstract

A growing number of investors and other stakeholders are demanding greater transparency about companies' sustainability performance. Countries around the world are responding with mandatory environmental, social and governance (ESG) disclosure regulations. On the one hand, this has led to a fragmented regulatory landscape. On the other hand, there is little empirical evidence on how these mandatory ESG disclosure requirements affect reporting companies. This paper addresses the gap through a three-part approach. First, it examines the evolution and status quo of global ESG regulation. Second, it provides a comprehensive literature review on the impact of such regulation on affected firms. Finally, it applies a difference-in-difference analysis to assess whether the EU's Non-Financial Disclosure Directive (NFRD), an unprecedented supranational ESG disclosure regime, has led to the adoption of more sustainable business practices. The results show a stronger increase in ESG scores for regulated EU companies compared with a control group of non-regulated US companies. This finding has practical relevance not only for the EU, but also for other legislators considering the introduction or expansion of ESG reporting requirements.

Keywords: disclosure; ESG regulation; NFRD; reporting; sustainability

1. Introduction

Sustainable investing has been on the rise in recent years, with global investments in sustainable assets exceeding USD 30 trillion in 2022 (GSIA, 2023, p. 10). To make well-informed investment decisions, investors (and other stakeholders) are increasingly demanding reliable and comprehensive information on companies' environmental, social, and governance (ESG) data (Amel-Zadeh & Serafeim, 2018, p. 2). However, voluntary ESG disclosure has often resulted in insufficient quality and quantity of reported data (H. B. Christensen et al., 2021, p. 1208). Recognizing this problem, regulators around the world are increasingly implementing mandatory ESG disclosure regimes to improve the comparability, consistency, and reliability of ESG reporting. As a result, more companies are being forced to provide detailed ESG information. (European Commission, 2014, p.2; Ioannou and Serafeim, 2017, p. 2; SEC, 2022b, pp. 28-29)

This regulatory shift has sparked intense public debate. Proponents argue that mandatory ESG reporting not only increases transparency for stakeholders, but also provides

significant benefits to the reporting companies themselves. These benefits range from enhancing corporate reputation and attracting talent to improving operational efficiency by identifying and mitigating risks and inefficiencies. (Ioannou & Serafeim, 2017, p. 5) However, critics point to several challenges associated with ESG reporting. The lack of standardized ESG metrics leads to diverse and potentially less comparable data across companies. Because ESG information is often qualitative in nature, the reporting relies on management's subjective assessment. Even quantitative ESG aspects are difficult to verify by external auditors, whose review is typically limited to an assessment of the company's policies and processes. In addition, companies often lack control and oversight over certain ESG disclosure requirements, such as those related to their supply chain. As a result, companies subject to ESG disclosure requirements may face significant operational and financial burdens associated with data collection and disclosure, as well as potential necessary changes to business processes. (H. B. Christensen et al., 2021, pp. 1227-1228; Ioannou and Serafeim, 2017, p. 5)

The impact of ESG disclosure on firms therefore remains uncertain and has become the subject of empirical research. Although *voluntary* ESG disclosure has been extensively studied in the literature, research on the effects of *mandatory* ESG disclosure requirements is still limited and inconclusive (H. B. Christensen et al., 2021, p. 1208; Mittelbach-Hörmanseder et al., 2021, p. 313). Studies provide mixed evidence on whether ESG disclosure requirements lead to changes in corporate reporting practices (e.g., Chauvey et al., 2015; Hummel and Rötzel, 2019; Ottenstein et al., 2022), their impact on firm value and financial performance (e.g., Y.-C. Chen et al., 2018; Grewal et al., 2019; Ioannou and Serafeim, 2017), and whether they are effective in encouraging companies to adopt more sustainable business practices (e.g., Downar et al., 2021; Fiechter et al., 2022; Tomar, 2023). With respect to the latter, research tends to focus on targeted disclosure requirements. This includes mandates aimed at specific industries, such as the US manufacturing and retail sectors (She, 2022), specific countries, such as China (Y.-C. Chen et al., 2018), or specific disclosure metrics, such as greenhouse gas (GHG) emissions (Bauckloh et al., 2023; Tomar, 2023). However, the extent to which these findings on targeted regulation can be generalized across different regulatory environments remains uncertain (Fiechter et al., 2022, p. 1500).

Given the global trend towards more comprehensive and widespread ESG disclosure requirements, it is critical to gain a deeper understanding of the implications of such frameworks, particularly in the context of the European Union (EU). The EU's introduction of the Non-Financial Disclosure Directive (NFRD) in 2014 marked a significant milestone, imposing broad ESG reporting requirements on approximately 11,000 companies across countries and industries (European Commission, 2014, p. 2; European Parliament, 2023). Based on the notion that greater transparency promotes greater accountability, the NFRD also aims to encourage companies to improve their ESG-related corporate practices (European Commission, 2014, p. 1). Initial studies suggest a positive impact of the NFRD on ESG performance, but these primarily focus on anticipatory effects (i.e. prior to its entry into force) or early impacts from the first reporting years (e.g., Cuomo et al., 2022; Fiechter et al., 2022), leaving a gap in evidence on the longer-term effects. Understanding the effectiveness of the NFRD in driving real change is particularly relevant with the introduction of the Corporate Sustainability Reporting Directive (CSRD), which will require more companies to provide more detailed disclosure from 2024, for reports published in 2025 (European Commission, 2023d).

Based on the identified gaps in the literature, the objective of my thesis is threefold. First, I provide a comprehensive overview of the development and current state of mandatory ESG disclosure regulations across the globe. This is crucial as the rapid evolution has led to an untransparent regulatory landscape, leading to the first research question: *What is the evolution and current state of mandatory ESG disclosure regulations across the globe?*

Second, I critically review the existing literature to assess the impact of these regulations on affected firms. The rapidly evolving nature of mandatory ESG regulation underscores the need for timely analyses, especially given that much of the prior literature reviews focus primarily on voluntary reporting (Haji et al., 2023, p. 178). This leads to the second research question: *What is the impact of mandatory ESG disclosure requirements on affected firms, according to current findings in the literature?*

Third, I aim to fill the research gap regarding the effectiveness of comprehensive ESG disclosure regimes in promoting the adoption of more sustainable corporate behavior. I empirically examine the impact of the NFRD on the ESG scores of affected EU firms relative to a control group of unregulated US firms, using a difference-in-difference (DID) design. The third question is therefore: *What is the impact of the NFRD on the ESG performance of affected firms?*

This thesis contributes to the literature in several meaningful ways. First, it shows that while global mandatory reporting requirements differ across categories (such as scope, disclosure content, and enforcement mechanisms), there is a trend toward greater convergence of regulations. Second, the literature review shows that ESG disclosure regulation seems to have a positive impact on the quantity and quality of disclosure, although there is also conflicting evidence. The literature on the financial impact of ESG disclosure regulation provides mixed evidence on both capital market and profitability effects. ESG disclosure requirements appear to be effective in encouraging companies to improve their ESG performance, although the evidence to date is mostly focused on targeted regulation. Third, the empirical analysis of the effects of the NFRD provides new insights into role of ESG reporting requirements in transitioning towards a more sustainable economy. I provide empirical evidence that the comprehensive disclosure directive positively affects firms' ESG behavior, thereby filling a critical gap in long-term impact studies. A thorough understanding of the different ESG disclosure requirements and their consequences on reporting firms is crucial for regulators, companies potentially affected by the mandates, and researchers seeking to study their impact.

The remainder of this paper is structured as follows: Chapter 2 provides an overview of global ESG disclosure regulations. Chapter 3 reviews the literature on the effects of mandatory ESG regulations, focusing on three areas: impact on ESG disclosure practices (Section 3.1), financial impact (Section 3.2), and impact on ESG-related firm performance (Section 3.3). Chapter 4 presents the empirical analysis of the impact of the NFRD on ESG performance. Chapter 5 concludes.

2. ESG disclosure regulations across the globe

In this chapter, I first provide background information on ESG disclosure, including its origins and a definition of ESG (Section 2.1). Second, I provide an overview of the main developments of mandatory ESG disclosure regimes in selected

regions around the world (Section 2.2). A summary of the regulations is also provided in Appendix 1. Third, I compare the different regulatory approaches across a number of dimensions (Section 2.3).

2.1. Background of ESG reporting

Voluntary sustainability reporting emerged in the 1960s in the US and Europe in response to a growing sense of social and environmental responsibility. In the 1980s, ethical funds in the UK and US adopted a negative screening investment approach, actively excluding companies in certain industries like tobacco and alcohol. In response to a major environmental disaster in the oil industry in the late 1980s, the first environmental reporting guidelines were developed. In the late 1990s, the Global Reporting Initiative (GRI) was launched with the aim of establishing comprehensive, credible and comparable corporate disclosure on economic, sustainability and social issues. The number of sustainability reports published on a voluntary basis has increased significantly as society demanded greater transparency and accountability. At the same time, investors have increasingly integrated sustainability factors into their investment decisions, further increasing the demand for comparable and reliable information. As social and environmental challenges become more pressing, concerns about the sufficiency of the voluntary approach have increased, leading to a growing number of jurisdictions around the world introducing mandatory reporting requirements. (Ioannou & Serafeim, 2017, pp. 6-7) While in the past mandatory regulations were mostly aimed at increasing environmental disclosure in specific sectors, there is a trend towards more comprehensive ESG disclosure regimes affecting companies across industries (Haji et al., 2023, p. 180).

ESG refers to the three pillars of sustainable corporate behavior, i.e., the activities companies undertake to be more economically and socially responsible, and their governance and control mechanisms. The environmental pillar includes issues related to climate change, toxic emissions, and waste, amongst others. Social issues include human rights and employee health and safety, and the governance pillar is related to issues regarding anti-corruption and anti-bribery measures, as well as auditing processes, management compensation, and disclosure practices, among others. (United Nations Global Compact, 2004, p. 6) *ESG disclosure* refers to the information that a company provides to the public on these issues (Ottenstein et al., 2022, p. 77).¹

¹ “ESG” is often used interchangeably with the terms “sustainability”, “corporate social responsibility” (CSR), and “non-financial” in related literature (Ioannou & Serafeim, 2017, p. 2). For consistency purposes, I use “ESG” throughout my thesis. In line with related literature, I also use the terms “disclosure” and “reporting” interchangeably, as I refer to information that is made available to the public (Ottenstein et al., 2022, p. 77).

2.2. Development of mandatory ESG disclosure regulations in selected regions

I limit the regulations in scope of my analysis to a few key criteria. First, I focus exclusively on ESG reporting requirements and exclude outcome-based regulations such as environmental protection laws. Second, I limit the analysis with respect to the target group of the regulations. ESG disclosure regulations can be divided into three categories: (1) disclosure regulations for corporations, (2) classification systems (i.e., taxonomies) for sustainable activities, and (3) fund disclosure requirements for financial market participants like asset managers or insurance companies (European Commission, 2023e). I focus on the first category. Third, I exclude voluntary reporting schemes and focus on mandatory requirements. This includes regulations that are implemented either on a fully mandatory basis or on a comply-or-explain basis. The latter gives companies the opportunity to provide explanations in the event of non-compliance with disclosure requirements (H. B. Christensen et al., 2021, p. 1228). The analysis includes both targeted, industry-specific regulations with limited disclosure requirements, as well as broad ESG disclosure regimes. A summary of the regulations is provided in Appendix 1.

2.2.1. Europe

This section first provides an overview of the evolution of ESG disclosure regulation in the EU, then discusses the main EU disclosure directives, and concludes with an overview of additional disclosure requirements implemented by selected European countries.

Development of ESG disclosure regulation in the EU

As shown in Figure 1, the EU first addressed ESG disclosure in its Directive 2003/51/EC, which emphasized the inclusion of environmental and social issues in firms’ financial reporting, on a voluntary basis (Agostini et al., 2022, p. 81; European Parliament and Council, 2003, p. 17). Since then, the EU has made several attempts to encourage voluntary ESG reporting, including through Directive 2013/34/EU (European Parliament and Council, 2013, p. 38). However, the voluntary provisions proved to be ineffective, resulting in less than 10% of large EU companies disclosing their ESG-related information. An important milestone towards mandatory ESG disclosure was the adoption of the Non-Financial Reporting Directive (Directive 2014/95/EU, NFRD) in 2014. The NFRD amends Directive 2013/34/EU by requiring large public interest corporations to include ESG information in their reports. (European Commission, 2014, pp. 1-2)

In 2015, the EU has committed to major international climate and sustainability goals by adopting the UN 2030 Agenda for Sustainable Development with its Sustainable

Development Goals (SDGs)², and the Paris Agreement³ on climate change (European Parliament and Council, 2020b, p. 13). In order to achieve the objectives of the UN 2030 Agenda and the Paris Agreement, the European Commission released its Action Plan: Financing Sustainable Growth (Action Plan) in March 2018. The plan proposes the creation of an EU sustainability taxonomy and the improvement of corporate ESG disclosure requirements. (European Commission, 2018a, 2018b, 2018c)

As part of the 2019 European Green Deal, with its overarching goal of making Europe climate neutral by 2050, the EU committed to enhancing its Action Plan, including the commitment to further increase ESG transparency for stakeholders (European Commission, 2018b, 2023d). In the same year, the Commission adopted the Sustainable Finance Disclosure Regulation (SFDR). The SFDR requires financial market participants (such as asset managers, insurance companies, pension funds, and investment firms) to disclose ESG-related information about their financial services and products, and entered into force in 2021. (European Commission, 2023g; European Parliament and Council, 2019)

The EU Taxonomy in 2020 introduced a new classification system aimed at creating a common understanding of sustainable economic activities (European Parliament and Council, 2020b). The Corporate Sustainability Reporting Directive (CSRD) was released in 2022 and extends the NFRD in terms of the scope of companies affected and the breadth of reporting requirements (European Parliament and Council, 2022). In the same year, the Commission has also adopted a proposal for the Corporate Sustainability Due Diligence Directive (CSDDD), which requires companies to mitigate their negative impacts on human rights and the environment throughout their supply chains (European Commission, 2022b).

Current ESG disclosure regulations in the EU

This section provides a more detailed analysis of the EU regulations that target ESG disclosure of corporates: the NFRD, the CSRD, and the EU Taxonomy.⁴

Non-Financial Disclosure Directive (NFRD)

The NFRD was adopted by the EU in 2014 to make ESG-related disclosures more transparent, consistent, and com-

² The UN 2030 Agenda is a global initiative promoting sustainable development through 17 Sustainable Development Goals (SDGs), covering areas such as poverty, health, education, gender equality, access to water and energy, and environmental sustainability. (United Nations, 2015, p. 14)

³ 195 countries have committed to the Paris Agreement, which is a legally binding international treaty aiming to strengthen global resilience to climate change, and to limit global warming to well below 2°C. (European Union, 2016)

⁴ Both the SFDR and the CSDDD are outside the scope of my analysis, as the former is aimed at financial market participants and the latter is not primarily considered a transparency tool (but rather a due diligence obligation). Although sustainability taxonomies are not specifically within scope, I do include the EU Taxonomy as it requires companies subject to the NFRD to disclose the proportion of their activities that is considered sustainable.

parable across companies in the European Union. It targets large public interest entities⁵, such as listed companies, banks, and insurance companies with more than 500 employees (Recital 14). (European Parliament and Council, 2014) Approximately 11,000 companies across the EU are affected (European Parliament, 2023).

Specifically, companies subject to the NFRD must report on environmental matters (e.g., environmental and health and safety impacts, energy use, greenhouse gas emissions, water use, and air pollution), social and employee-related matters (e.g., working conditions and measures to ensure gender equality), protection of human rights, and anti-corruption and anti-bribery matters (e.g., mechanisms in place to prevent such matters) (Recital 7). In relation to these matters, companies are required to provide details of their policies, outcomes and risks, as well as their due diligence processes, including of their supply chain, where relevant (Recital 6). The information must be presented in such a way that the company's development, performance, position, and impact in relation to these issues are sufficiently clear (Article 19a(1)). In addition, certain large companies must disclose their policy on board diversity (i.e., gender, age, and background of board members) (Recital 19). EU Member States had to transpose the NFRD into national law until December 2016.⁶ 2018 was the first year companies subject to the law had to disclose ESG information in line with the NFRD for the 2017 financial year (Article 4). (European Parliament and Council, 2014)

However, several shortcomings led to a revision of the NFRD and ultimately to the adoption of the new CSRD (European Parliament, 2023). The shortcomings include a lack of mandatory compliance with reporting standards and formats, and a lack of external assurance requirements on disclosed content. In addition, companies can justify their non-compliance with the NFRD's disclosure requirements by providing reasoned explanations on a comply-or-explain basis. Finally, although companies are required to report according to the double materiality principle⁷, the directive lacks a clear definition of this. (EPRS, 2021, pp. 2-4)

Corporate Sustainability Reporting Directive (CSRD)

The CSRD entered into force in January 2023 and introduced stricter reporting requirements. The scope of com-

⁵ As defined in Article 3(4) of Directive 2013/34/EU, companies qualify as large undertakings if they exceed at least two of the three thresholds: a net turnover of EUR 40 million, a balance sheet total of EUR 20 million, and/or an average number of 250 employees during the year. (European Parliament and Council, 2013, p. 30)

⁶ The NFRD provides EU member states some freedom in the transposition into their respective national law. For example, France, Italy, and Spain adopted mandatory external assurance requirements, while most of the other countries have not (Accountancy Europe, 2020, p. 2). A comprehensive overview of the differences in requirements between member states is provided by CSR Europe and GRI (2017).

⁷ Double materiality refers to the underlying methodology for assessing material (i.e., relevant) information to be disclosed, which considers both financial and impact perspectives. See Section 2.3 for a detailed discussion.



Figure 1: Development of ESG disclosure regulation in the EU
(source: own representation, based on Blaschke et al., 2023, p. 14)

panies affected was extended to all large companies⁸ (regardless of capital market orientation) and all listed Small and Medium Enterprises (SMEs), except micro-enterprises (Recital 17). The CSRD also covers non-EU companies generating a net turnover of over EUR 150 million in the EU, and with a subsidiary in the EU being a large company or a listed SME, or with a branch generating a turnover of over EUR 40 million, are subject to the CSRD (Recital 20). (European Parliament and Council, 2022, pp. 19-20) Under the CSRD, approximately 50,000 companies will be affected by mandatory ESG disclosure requirements, a significant expansion from the 11,000 companies subject to the NFRD (European Parliament, 2023).

The CSRD requires reporting in accordance with the European Sustainability Reporting Standards (ESRS), addressing the lack of comparability and reliability of reporting under the NFRD. The ESRS were developed by the European Financial Reporting Advisory Group (EFRAG) and endorsed by the EU in July 2023 (European Commission, 2023f). The framework consists of twelve standards: two overarching standards (ESRS 1 and ESRS 2) outline the fundamental reporting principles and obligations (European Commission, 2023a, p. 3). These are followed by ten topical disclosure standards, which cover the environmental (climate change, pollution, water and marine resources, biodiversity and ecosystems, resource use and circular economy), social (own workforce, workers in the value chain, affected communities, consumers, and end users), and governance pillar (business conduct, including animal welfare, political engagement, and corruption and bribery) (European Commission, 2023a, pp. 24-26). With respect to these topics, the ESRS require disclosure in the areas of governance, strategy, impact, risk and opportunity management, as well as certain quantitative metrics and targets. In addition, the ESRS define certain industry-specific standards that only apply to companies in specific industries. (European Commission, 2023a, pp. 3-4)

Companies must report on all topics that are material (i.e., relevant) to them, and otherwise provide a detailed explanation as to why the specific topic is not material to them.⁹

⁸ Large companies are defined by Article 3(4) of Directive 2013/34/EU (see footnote 5).

⁹ For a detailed explanation of the materiality assessment process, please

In assessing the materiality of topics, companies are required to adopt a double-materiality approach. This means that issues must be evaluated in terms of both their *financial materiality*, which refers to the company's exposure to climate change risks and opportunities, and their *impact materiality*, which refers to the company's impact on the environment and society. A topic must be disclosed if it is material to either or both dimensions. (European Commission, 2023a, pp. 7-8) The materiality assessment is also subject to mandatory verification by an external assurance provider (European Commission, 2023f). The CSRD will be phased in, with the first group (consisting of companies already subject to the NFRD) required to comply from financial year 2024 for reports published in 2025 (European Commission, 2023f; European Parliament, 2023).¹⁰ The directive must be transposed into national law by all EU member states until July 2024 (European Parliament and Council, 2022).

EU Taxonomy Regulation

The EU Taxonomy Regulation (Regulation (EU) 2020/852) is a classification system that aims to provide a common understanding and definition of sustainable economic activities. To be considered sustainable, activities must contribute to at least one of six environmental objectives: (1) climate change mitigation, (2) climate change adaptation, (3) sustainable use and protection of water and marine resources, (4) transition to a circular economy, (5) pollution prevention and control, and (6) protection and restoration of biodiversity and ecosystems. In addition, the activity must not harm any of the other objectives. The Taxonomy became effective in July 2020. (European Parliament and Council, 2020a)

Since January 2022, the Disclosures Delegated Act complements the EU Taxonomy by requiring companies subject to the NFRD to disclose the proportion of their taxonomy-

refer to Appendix E of the ESRS. (European Commission, 2023a, p. 33)

¹⁰ Other large companies not currently covered by the NFRD will have to report in accordance with the CSRD from 2025 for reports published in 2026. Listed SMEs will have to report from 2026 for reports published in 2027 (with a reduced set of standards), with the possibility to opt out for a further two years. Non-EU companies with significant activities in the EU that fall within the scope of the CSRD will have to comply from 2028 for reports published in 2029, and will also be subject to separate reporting standards. (European Commission, 2023f; European Parliament, 2023)

aligned economic activities that contribute to their revenue, capital expenditure (CapEx), and operating expenditure (OpEx). (European Parliament and Council, 2021)

Additional regulations of selected European countries

This section gives an overview of individual reporting requirements enacted by certain European countries, focusing on Germany, France, and the UK.

Germany

Prior to the NFRD, Germany did not have a comprehensive mandatory ESG disclosure requirement. Germany transposed the NFRD into national law as the *Gesetz zur Stärkung der nichtfinanziellen Berichterstattung der Unternehmen in ihren Lage- und Konzernlageberichten (CSR-Richtlinie-Umsetzungsgesetz)* in April 2017 (Bundestag, 2017). Compared to other EU member states such as Italy and France, Germany has taken a conservative implementation approach, as reflected by the adoption of the minimum reporting requirements prescribed by the NFRD (Jeffery et al., 2017, pp. 2-3). In an analysis of countries' positions during the Commission's negotiations regarding the enactment of the NFRD, Kinderman (2020, p. 680) finds that Germany was the strongest opponent of the directive.

In July 2021, Germany passed the Act on Corporate Due Diligence in Supply Chains (*Lieferkettensorgfaltspflichtengesetz*), which became effective in January 2023. To ensure corporate responsibility for respecting human rights and protecting the environment in global supply chains, the Act implements specific due diligence obligations with respect to the company itself and its contractual partners and direct and indirect suppliers (Paragraph 3). It applies to companies with more than 3,000 employees in Germany from 2023 and to companies with more than 1,000 employees in Germany from 2024 (Paragraph 1). According to Paragraph 10(2) of the Act, affected companies are required to annually publish a report on the fulfillment of their due diligence obligations on their corporate website. The disclosure must include information on human rights and environmental risks and violations, measures taken to fulfill the obligations, and an assessment of their effectiveness. (Bundestag, 2021)

France

France already introduced mandatory disclosure of a social report (*Bilan Social*) in 1977. The law required companies with more than 300 employees to report on 134 social indicators related to their employment activities (Antal and Sobczak, 2007, p. 15; Government of France, 1977).

In 2001, the Law on New Economic Regulations (*Loi Nouvelle Régulation Economique*, NRE) was introduced, requiring listed French companies to disclose social and environmental impacts in their annual reports (Government of France, 2001; Ministère des Affaires Etrangères France, 2012, p. 1). With this regulation, France became one of the first coun-

tries in the world to require social and environmental disclosure (Chauvey et al., 2015, p. 789). The NRE was amended by the *Loi Grenelle I* and *Loi Grenelle II* in 2009 and 2010, respectively (Government of France, 2009, 2010). All companies with more than 500 employees were required to disclose ESG-related information on a comply-or-explain basis. The regulations expand the disclosure requirements to more than 40 social and environmental topics, which are subject to mandatory independent verification. (Ministère des Affaires Etrangères France, 2012, pp. 4-5)

In March 2017, the French Law on Corporate Duty of Vigilance (*Devoir de Vigilance*) came into force, aiming to prevent human rights violations, safeguard public health and safety, and protect the environment. The law requires French companies with more than 5,000 employees in France and international companies with more than 10,000 employees in France to disclose a plan outlining their due diligence processes with respect to specific ESG-related matters. Companies may face penalties for not providing a vigilance plan. (Business & Human Rights Resource Centre, 2017; Government of France, 2017a)

France transposed the NFRD into national law as the Ruling No. 2017-1180 of July 19, 2017 (Aureli, Salvatori, and Magnaghi, 2020, p. 9; Government of France, 2017b). The ESG-related disclosure requirements go beyond those mandated by the NFRD, including, for example, mandatory disclosure of employee diversity, impact on biodiversity, and measures taken in relation to the circular economy (Aureli, Salvatori, & Magnaghi, 2020, p. 15).

UK

Since 2006, quoted companies¹¹ in the UK are required to publish a business review as part of the Companies Act 2006. Section 417 of the Act mandates the disclosure of a companies' environmental impact, as well as their impact on employees and the society. (Parliament of the United Kingdom, 2006b)

In October 2013, the Companies Act (Strategic Report and Directors' Report) Regulations 2013 became effective. The amendment to the Companies Act 2006 replaces the business review (i.e., Section 417 of the Companies Act 2006) with the requirement for large and medium-sized companies to disclose a more comprehensive strategic report.¹² (Jackson et al., 2020, p. 339) Specifically, the Act mandates the disclosure of non-financial Key Performance

¹¹ A UK quoted company is defined as a company incorporated in the UK whose equity share capital is either listed on the Main Market of the London Stock Exchange, is officially listed in an EEA state, or is admitted to trading on the Nasdaq or the New York Stock Exchange. (Parliament of the United Kingdom, 2006a, p. 11, section 385(2))

¹² Small companies are exempt from the obligation to publish a strategic report. According to section 382 of the Companies Act, a company qualifies as small by meeting at least two of the three conditions: maximum turnover of GBP 10.2 million, maximum balance sheet total GBP 5.1 million, or a maximum of 50 employees. (Parliament of the United Kingdom, 2006a, p. 3)

Indicators (KPIs) concerning environmental and employee matters (Section 414C Subsection 4(b)). In addition, it requires disclosure of principal risks and uncertainties, which also include ESG issues (Section 414C Subsection 2(b)). For UK quoted companies, the strategic report must include information regarding environmental issues (including the company's impact on the environment), its employees, as well as social, community and human rights matters (Section 414C Subsection 7(b)). Quoted companies are also mandated to include information regarding gender diversity on board level and in senior management positions as well as the entire company (Section 414C Subsection 8(c)). (Parliament of the United Kingdom, 2013, pp. 2-4) As the first country worldwide (Krueger, 2015, p. 3), the UK requires mandatory Scope 1 and 2 GHG emissions¹³ disclosure for quoted companies (Part 7) (Parliament of the United Kingdom, 2013, p. 5).

The NFRD was transposed into national law through the addition of Section 414CA and 414CB to the Companies Act 2006. This expands the scope of companies subject to mandatory ESG disclosure, and imposes additional disclosure requirements regarding anti-corruption and anti-bribery matters. (Parliament of the United Kingdom, 2016)

In 2022, the Companies Act 2006 was further amended by the Companies (Strategic Report) (Climate-related Financial Disclosure) Regulations 2022 (Parliament of the United Kingdom, 2022). Economically significant companies¹⁴ within the scope of the regulation are required to disclose their risks and opportunities related to climate change in alignment with the Task Force on Climate-related Financial Disclosures (TCFD) standards. (UK Department for Business, Energy and Industrial Strategy, 2022, pp. 5-6; UK Government, 2023b)

Depending on the type of company, certain entities are also required to disclose additional information, i.e., regarding modern slavery (for companies with more than GBP 36 million turnover) or the gender pay gap (for companies with more than 250 employees). (UK Government, 2023a, 2023b)

Other European countries

Other countries that have introduced ESG disclosure regulations prior to the enactment of the NFRD are Denmark, Sweden, and Spain (European Commission, 2014, p. 1). Denmark has mandated large companies to disclose ESG-related information together with their annual statements

since 2008. Sweden and Spain have introduced ESG disclosure requirements in 2007 and 2011, respectively. However, the Swedish regulation covered only state-owned enterprises, and the Spanish requirement was implemented on a recommendation basis, and not on a mandatory basis. (Ioannou and Serafeim, 2017, p. 8; Jackson et al., 2020, p. 328)

2.2.2. United States

The US lags behind the EU in terms of mandatory ESG disclosure regulation (Cicchello et al., 2023, p. 1122). While existing regulations require companies to include environmental issues in financial, litigation, and risk disclosures, there are no mandatory requirements for broad ESG disclosures to date. Nonetheless, the Securities and Exchange Commission (SEC) has issued a proposed rule on climate disclosure in 2022, which would introduce extensive ESG disclosure requirements if adopted (SEC, 2022a). Individual states, especially California, are also progressing to advance their own ESG disclosure regime.

Regulations at the federal level

In 1971, the SEC has first published a proposal for registered companies to disclose how complying with federal, state, and local environmental laws impacts them financially. In 1982, the SEC mandated the disclosure of business costs (including litigation costs) resulting from compliance with environmental protection laws. In addition, companies are required to report on climate change-related matters as part of other, more general SEC disclosure requirements. (SEC, 2022b, pp. 15-16) In 2010, the SEC issued the Interpretive Guidance on Disclosures Related to Climate Change, to assist companies in complying with these disclosure regulations related to climate change (SEC, 2010, p. 3). Specifically, companies must consider climate change issues that could have a material impact on their financial condition in their SEC filings¹⁵ (SEC, 2010, pp. 12-15). However, these obligations have not resulted in sufficiently comparable and consistent ESG disclosures across reporting companies, as demanded by investors (SEC, 2022b, p. 296).

In an effort to strengthen the 2010 voluntary guidance, the SEC proposed the Enhancement and Standardization of Climate-Related Disclosures for Investors in March 2022, which would make climate-related disclosures mandatory. The proposed rules would require SEC registrants to disclose information including on climate risks and their (potential) impact on the company, the governance and management processes in place to address these risks, GHG emissions

¹³ Scope 1 are direct emissions from a company's own operations, Scope 2 refers to indirect emissions from the generation of purchased energy, and Scope 3 includes all other indirect emissions across a company's value chain, both upstream and downstream. (GHG Protocol, 2023)

¹⁴ The following companies qualify as being "economically significant": listed companies, banks or insurers with at least 500 employees, UK-based AIM companies with at least 500 employees; LLPs with at least 500 employees and a turnover of over GBP 500 million, and non-listed companies with at least 500 employees and a turnover of over GBP 500 million. (UK Government, 2023b)

¹⁵ Particularly, disclosure requirements related to climate change include reporting on the impact of climate change related regulation and international accords, such as the Kyoto Protocol. Registrants are furthermore required to disclose any indirect risks or opportunities of regulation or business trends, such as lower demand for high GHG-emitting goods, or a rising demand for alternative energy. Companies must also include physical consequences of climate change in their disclosures, such as the risks and effects of severe weather conditions and other climate related events. (SEC, 2010, pp. 22-27)

(Scope 1, 2 and, if material, Scope 3 emissions), specific climate-related financial figures, and climate-related targets with a transition plan. The proposed rules are based on commonly used disclosure frameworks (such as the TCFD). Disclosure would be required as part of companies' registration statements and periodic reports. In addition, accelerated filers and large accelerated filers¹⁶ would be required to obtain external assurance on at least their Scope 1 and Scope 2 GHG emissions. The SEC is proposing initial disclosure under the new rule in 2024 reports for fiscal year 2023 for large accelerated filers, and in 2025 reports for fiscal year 2024 for accelerated filers (excluding Scope 3 emissions). (SEC, 2022a, 2022b, pp. 41-43) The rulemaking is currently pending. Opponents argue that the proposed requirements are beyond the SEC's authority. The scope of the reporting requirements is also likely to be compromised, as the current version of the proposal is expected to face legal challenges. (Mcgowan, 2023)

In addition to these developments towards a broad ESG disclosure regime, the SEC has also adopted several more targeted disclosure requirements. For example, Section 1502 of the Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank Act) requires SEC-registered companies to report on the use of conflict minerals (i.e., tantalum, tin, gold, and tungsten) from the Democratic Republic of the Congo (DRC) or neighboring countries since 2012 (SEC, 2012a, 2012b). Section 1503 requires mine owners to disclose health and safety violations, and mining-related fatalities, among others (SEC, 2012c). Section 1504 requires listed oil and gas companies to disclose payments made to the US or foreign governments (SEC, 2020; US Congress, 2010, p. 2220), aiming to improve governance in extraction countries (Healy & Serafeim, 2020, p. 112).

The SEC has also recently passed disclosure requirements related to diversity and cybersecurity. Since August 2021, Nasdaq-listed companies must have at least two diverse board members (on a comply-or-explain basis). In addition, racial identity, LGBTQ+ status and voluntary self-disclosed gender identity of all board members must be disclosed. (Nasdaq, 2023; SEC, 2021, pp. 3-4) Since September 2023, public companies are required to report any material cybersecurity incident promptly, and to disclose information on cybersecurity risk management, strategy, and governance annually (SEC, 2023)

The Environmental Protection Agency (EPA) has released regulatory programs requiring the disclosure of specific environmental information. Under the Greenhouse Gas Reporting Program (GHGRP), major emitters in the US are required to disclose their GHG emissions since 2010 (US EPA, 2009). The information is verified and published by the EPA on their website (US EPA, 2023a). The Toxics Release Inventory (TRI) requires larger facilities in specific industries (including metal mining and electric power generation) to an-

nually disclose their emission of certain toxic chemicals (US EPA, 2023b).

Regulations at the state level

ESG disclosure regulations at the state level are mostly specific to certain disclosure items and aim to indirectly nudge companies towards the adoption of more sustainable behavior (Ho, 2017, p. 326). For example, at least 17 US states have implemented additional GHG emissions disclosure requirements, mostly regarding Scope 1 emissions (SEC, 2022b, pp. 298-299).

In California, the California Transparency in Supply Chains Act (CTSCA) of 2010 requires large retail sellers and manufacturers doing business in California to disclose their due diligence processes and efforts to ban human rights abuses from their supply chains since 2012 (California Government, 2010).

In October 2023, the California Senate enacted two bills which introduce extensive climate-related disclosure requirements for large public and private firms operating in California, becoming effective in 2026 (Engler, 2023). Senate Bill (SB) 253 (The Climate Corporate Data Accountability Act) requires companies with more than USD 1 billion revenues doing business in California to disclose Scope 1, 2, and 3 GHG emissions, including independent assurance requirements (California Government, 2023a). SB 261 (Greenhouse gases: climate-related financial risk) requires companies doing business in California with annual revenues over USD 500 million to report their climate-related financial risk, in line with TCFD reporting recommendations (California Government, 2023b). The California laws are even more stringent than the SEC's proposed rules. First, they apply to large private companies, whereas the SEC proposal applies only to SEC-registered companies (Engler, 2023). Second, they require Scope 3 emissions reporting in all cases, while the SEC proposal exempts some companies (SEC, 2022b, p. 43).

2.2.3. APAC region

In Asia Pacific, the analysis focuses on ESG disclosure developments in India, China, Hong Kong, Japan, Malaysia, and Australia.

India

In India, the two main sources of ESG regulation are the Ministry of Corporate Affairs (MCA) and the Securities and Exchange Board of India (SEBI). Mandatory ESG reporting was introduced by the SEBI in 2012, requiring the top 100 companies by market capitalization to disclose their ESG-related performance in a Business Responsibility Report (BRR) as part of their annual reports (SEBI, 2012). The scope of affected companies has been extended to the top 500 companies in 2015 and to the top 1,000 companies in 2017 (PwC India, 2021, p. 5).

In 2013, the MCA enacted the Indian Companies Act, which came into force in April 2014 (MCA, 2013, p. 14).

¹⁶ An accelerated filer has a public float of USD 75 million or more, a large accelerated filer has a public float of USD 700 million or more. For a detailed definition, see footnote 122 and 123 of SEC (2022b, p. 41).

Section 135 of the Act requires companies above a certain size threshold¹⁷ to establish a CSR committee and disclose their ESG-related policies and expenditures, as well as the composition of their CSR committee. In contrast to the other disclosure regulations discussed, it also requires companies to spend at least 2% of their average net profits over three years on ESG activities annually on a comply-or-explain basis. (MCA, 2013, p. 87) Under Section 134(3)(m), companies must disclose information on energy conservation as part of their annual financial statements (MCA, 2013, p. 85). The Act is the first mandatory regulation in the world to require ESG spending in addition to ESG disclosure (Manchiraju & Rajgopal, 2017, p. 1269).

In 2021, SEBI replaced the BRR with the Business Responsibility and Sustainability Report (BRSR) and introduced new ESG disclosure requirements for the top 1,000 listed companies by market capitalization, effective from the financial year 2022-23 (SEBI, 2021a). Under the BRSR framework, companies are required to provide disclosures in three categories, namely general disclosures, management and process disclosures and principle-based performance disclosures. In the latter category, companies must provide quantitative measures along the nine principles of the National Guidelines on Responsible Business Conduct (NGRBC). These include issues related to human rights, employee welfare, resource use, GHG emissions (Scope 1 and 2), and anti-corruption and anti-bribery measures. (SEBI, 2021b, 2021c)

In July 2023, SEBI introduced independent assurance requirements for certain ESG metrics (BRSR Core). In addition, the top 250 companies by market capitalization are now subject to new value chain disclosure requirements for (on a comply-or-explain basis). (SEBI, 2023) Despite this general movement towards greater ESG transparency through mandatory disclosure requirements in India, actual compliance by companies is questionable due to weak enforcement in the country (Dharmapala & Khanna, 2018, p. 102).

China

In China, the scope of mandatory ESG regulation is still relatively limited to a subset of publicly listed companies and companies operating in highly polluting industries. ESG disclosure mandates were first introduced in 2008 by the Shenzhen Stock Exchange (SZSE) and the Shanghai Stock Exchange (SSE), both government-owned and overseen by the China Securities Regulatory Commission (CSRC). (Y.-C. Chen et al., 2018, p. 171)

The SZSE made reporting mandatory for companies listed on its Shenzhen 100 Index in December 2008. Article 11 of the Notice on the Preparation of 2008 Annual Reports of Listed Companies mandates ESG reporting in line with the

Requirements for the Disclosure of Corporate Social Responsibility (as stated in Appendix 3 of the Notice). The regulation requires reports to be approved by the board of directors and published as a separate report accompanying the annual report. The reports must include at least the following: First, it must provide an overview of the company's purpose in fulfilling its corporate social responsibility. Second, it must provide information on how it protects the interests of shareholders and creditors and the rights of employees, how it protects suppliers, customers, and consumers, how it preserves the environment and promotes sustainable development, and how it engages in public relations and social welfare initiatives. Third, it must disclose issues related to the company's fulfillment of its social responsibilities and specific plans and actions for improvement. (SZSE, 2008)

Similarly, the SSE issued the Notice of Doing a Better Job for Disclosing 2008 Annual Reports, which requires listed companies in its Corporate Governance Sector, companies with overseas listed shares, and financial companies to include ESG reports in their annual reports from 2008 (SSE, 2008, Article 10). The disclosure requirements are similar to those introduced by the SZSE (Y.-C. Chen et al., 2018, p. 172). Neither regulation requires mandatory assurance of reports, but companies face penalties in the form of delisting and public shaming for non-compliance (Y.-C. Chen et al., 2018, p. 171). Prior to these disclosure regulations, ESG reporting was rather uncommon for Chinese companies (Ioannou & Serafeim, 2017, p. 9).

Also in 2008, the State-owned Assets Supervision and Administration Commission of the State Council (SASAC) released the Notification on the Issuance of the Guideline on Fulfilling Social Responsibility by Central Enterprises. Article 18 of the notice mandates central state-owned enterprises to disclose ESG reports. (SASAC, 2008)

Besides these broader ESG disclosure regulations, several laws specifically target major polluting companies. In 2014, the Chinese government enacted the Environmental Protection Law. Article 55 of the law requires major emitters to disclose detailed information on their pollutant discharge. (Chinese Government, 2014) Since 2021, the CSRC requires listed key polluting companies to disclose specific environmental information regarding pollution and environmental discharge (CSRC, 2021; OECD, 2023). In the same year, the Ministry of Ecology and Environment (MEE) enacted the Administrative Measures for Legal Disclosure of Corporate Environmental Information. These measures require all key pollutant discharge units and listed companies with a history of environmental violations (as defined in Chapter 2 of the regulation) to disclose various environmental information, including information on emissions of pollutants, carbon emissions, and environmental emergency plans and measures (Chapter 3, Article 12). (MEE, 2021a)

Looking forward, the Plan for the Reform of the Legal Disclosure System of Environmental Information, released by the MEE in 2021, outlines the intention to establish a comprehensive disclosure system regarding environmental information by 2025 (MEE, 2021b).

¹⁷ The regulation applies to companies fulfilling at least one of the three criteria: (1) net worth of at least 5 billion Indian Rupees (about USD 60 million), (2) turnover of at least 10 billion Rupees (about USD 120 million), or (3) net profit of at least 50 million Rupees (about USD 0.6 million).

Hong Kong

The ESG Reporting Guide of the Hong Kong Stock Exchange (HKEX) requires all listed companies to disclose specific information regarding different ESG topics. Since its first introduction in 2013, the ESG Reporting Guide has continuously been revised, and its latest amendments came into effect in July 2020. (HKEX, 2021, p. 2) Part B of the current Guide outlines the mandatory ESG disclosure requirements, and Part C lists the provisions that must be disclosed on a comply-or-explain basis. Part B includes disclosure requirements on the governance structure (e.g., oversight of ESG issues, ESG strategy, and an assessment of progress against ESG goals and targets). In addition, the mandatory requirements include disclosure of the materiality approach and the quantitative measures used. (HKEX, 2019, p. 4) In Part C, environmental disclosure provisions include information on emissions (e.g., Scope 1 and Scope 2 emissions and waste disposal), use of resources (e.g., efficient use of water, energy, and other raw materials), environment and natural resources (e.g., measures to reduce environmental impact), and climate change (e.g., identification and mitigation strategies of climate issues). (HKEX, 2019, pp. 5-7) The social disclosure provisions include information on employment, health and safety, development and training, labor standards, supply chain management, product responsibility, anti-corruption, and community investment. (HKEX, 2019, pp. 7-11) Unlike the mainland Chinese stock exchanges, HKEX's ESG reporting requirements apply to all listed companies (and not just a subset of listed companies above a certain size threshold).

Japan

The Tokyo Stock Exchange (TSE) expanded its Corporate Governance Code in June 2021 to include new disclosure requirements for companies listed on the Prime Market (Japan Exchange Group, 2021). These include disclosure of companies' sustainability initiatives and investments in human capital, as well as the climate-related risks and opportunities in line with the TCFD recommendations (TSE, 2021, pp. 14-15). It also requires social disclosure on diversity in senior management and human resource development policies to promote diversity (TSE, 2021, p. 11).

Since 2021, the Financial Services Agency (FSA) requires all listed companies in Japan to disclose their sustainability-related policies and initiatives, including information on governance and risk management. Disclosure of sustainability-related strategy, metrics and targets (including Scope 1 and 2 GHG emissions) are required on a comply-or-explain basis. The requirements include diversity disclosures. (FSA, 2023; Lawless et al., 2023)

In terms of targeted regulation, high energy consuming factories are required to disclose their annual energy consumption under the Act on the Rational Use of Energy (Article 15(1)) (Government of Japan, 1979). In addition, large emitting companies are required to disclose their GHG emis-

sions to the government under the Mandatory Greenhouse Gas Accounting and Reporting System since 2006 (Japan Ministry of the Environment, 2006, 2021).

Malaysia

Bursa Malaysia, the country's stock exchange, requires listed companies to disclose ESG initiatives and activities in their annual reports since 2007, on a comply-or-explain basis (Bursa Malaysia, 2006, p. 8). Since 2015, companies must include a sustainability statement in their annual reports, including information on material risks and opportunities related to economic, environmental, and social matters¹⁸ (Bursa Malaysia, 2015c, p. 1; 2022, p. 8). Since 2023, the regulation requires disclosure of specific sustainability matters, including risks and opportunities related to anti-corruption, diversity, energy use, employee health and safety, supply chain management, data security, water and waste management, and emissions management (including Scope 1, 2, and 3 emissions) (Bursa Malaysia, 2022, pp. 72-73). Companies listed on the Main Market must report in accordance with the TCFD guidelines. In addition, companies must indicate whether their report has been subject to internal or independent assurance. (Bursa Malaysia, 2022, p. 8)

Australia

In 2023, the Australian Government released new consultation papers that initiate the transition from a voluntary to a mandatory ESG disclosure regime. The proposed rules are closely aligned with the International Sustainability Standards Board (ISSB) framework (Australian Government, 2023, p. 10). Disclosure requirements include information on material climate-related risks and opportunities, as well as the identification, assessment and mitigation strategies in place. In addition, disclosure of certain ESG metrics and targets (including Scopes 1, 2 and 3 emissions) would be mandatory. (Australian Government, 2023, pp. 12-17) The proposal also includes external assurance requirements (Australian Government, 2023, pp. 22) and application of financial materiality principles regarding risks and opportunities assessment (Australian Government, 2023, pp. 11-12). Large companies would be required to comply from the 2024 financial year (for reports filed in 2025) and medium and small companies from 2025 and 2026, respectively¹⁹ (Australian Government, 2023, pp. 8-9). Importantly, the new

¹⁸ As defined in Paragraph 29, Part A of Appendix 9C, of the Main Market Listing Requirements, and Paragraph 30 of Appendix 9C of the ACE Market Listing Requirements. (Bursa Malaysia, 2015a, 2015b)

¹⁹ Large enterprises fulfil two of the three criteria: more than 500 employees, over AUD 500 million revenues or over AUD 1 billion assets. Medium-sized companies fulfil two of the three size thresholds: more than 250 employees, over AUD 200 million revenue or AUD 500 million assets. Small companies fulfil two of the three thresholds: over 100 employees, over AUD 50 million revenues or over AUD 25 million assets. (Australian Government, 2023, pp. 8-9)

regulation would apply to both public and private companies meeting the size thresholds that are currently required to file financial reports under the Corporations Act (Australian Government, 2023, p. 6).

Besides the current development towards a broad mandatory ESG disclosure regime, some targeted reporting regulations exist already today in Australia. The National Greenhouse and Energy Reporting (NGER) Act 2007 mandates disclosure of GHG emissions, energy production and energy consumption for the largest emitting entities (as defined by Section 13 of the Act) (Australian Government, 2007, p. 42). Under the Modern Slavery Act 2018, large companies incorporated or doing business in Australia (as defined by Section 5 of the Act) are required to disclose their modern slavery risks in their operations and their supply chains, as well as their processes and mechanisms to address these risks (Part 2 of the Act) (Australian Government, 2018, p. 10).

2.2.4. Latin America

Although the implementation of mandatory ESG disclosure rules has been comparatively slow in Latin America, several jurisdictions have started to move towards a mandatory ESG disclosure regime.

Brazil

Since January 2023, the Brazilian Securities Commission (CVM) requires listed companies to disclose ESG-related information as part of its Resolution No. 59 (RCVM 59). On a comply-or-explain basis, Item 1.9 includes disclosure requirements on the location of ESG disclosure (i.e. annual report or other type of document), the methodologies used to prepare the report, the use of ESG KPIs and the materiality assessment, the status of independent verification, whether GHG emissions are measured (and, if so, their scope) and whether the TCFD recommendations (or other ESG standard) are taken into account (CVM, 2021, pp. 15-16). In addition, companies must disclose the environmental and social risks they face and the strategies and policies in place to mitigate them (CVM, 2021, pp. 27-28). The requirements also include information on diversity in management (CVM, 2021, pp. 38-39) and among employees (CVM, 2021, p. 60). However, RCVM 59 lacks a requirement to report on specific ESG issues and a proper enforcement mechanism, such as the imposition of penalties for non-compliance (Lavigne et al., 2021).

From 2026, listed companies in Brazil will be subject to more comprehensive ESG disclosure requirements. The CVM has issued Resolution No. 193 (RCVM 193) in October 2023, which requires ESG disclosure in line with the ISSB reporting standards and also includes mandatory reasonable assurance requirements. (CVM, 2023)

In terms of targeted disclosure regulations, since 2006, energy companies are required to publish an annual sustainability report, in accordance with the Aneel Guidelines for Annual Sustainability Report – Despacho 3034/2006. (UNEP, 2019, p. 17)

Chile

As part of the new climate change strategy, the Chilean Financial Markets Commission (CMF) announced in September 2020 its intention to increase ESG transparency, including through the implementation of stricter disclosure rules (CMF, 2020). In November 2021, the CMF introduced General Rule No. 461, which requires listed companies as well as financial institutions to include sustainability and governance matters in their annual reporting. The requirements include disclosure of climate-related risks and opportunities and respective mitigation measures, in line with international reporting guidelines like the TCFD (Paragraph 3.6). Additionally, companies are required to disclose industry-specific ESG metrics in line with the SASB reporting requirements, on a comply-or-explain basis (Paragraph 8.2). (CMF, 2021, 2022)

Argentina

Argentina does not have broad ESG disclosure requirements established. However, Law No. 2594 requires national and international companies in Buenos Aires with more than 300 employees to disclose an annual sustainability report (*Balance de Responsabilidad Social y Ambiental*) since 2008, in line with GRI reporting standards. (Government of Buenos Aires, 2007)

2.2.5. Other regions

This section extends the analysis of ESG disclosure requirements to regions not previously covered, focusing on Canada and South Africa.

Canada

To date, Canada does not have comprehensive mandatory ESG disclosure requirements. However, National Instrument 51-102 Continuous Disclosure Obligations requires public companies to disclose in their annual reports environmental and health risks material to their business (Item 5.2), as well as social and environmental policies (Item 5.1(4)), and the financial and operational impact of environmental regulations (Item 5.1(1)) since 2004 (CSA, 2004, pp. 3512-3513). In addition, there is a move towards more explicit mandatory ESG disclosure regulation. The Canadian Securities Administrators (CSA) has published the proposed National Instrument 51-107 Disclosure of Climate-related Matters in October 2021, which would mandate ESG reporting for listed companies in line with the TCFD recommendations. (CSA, 2021, p. 6)

With regards to GHG emissions reporting, Canada's largest emitters²⁰ are required to disclose their emissions as part of the Greenhouse Gas Reporting Program (GHGRP),

²⁰ Facilities with more than 10,000 tonnes of GHG emissions annually.

which was established under Section 46 of the Canadian Environmental Protection Act (CEPA) in 2004. (Government of Canada, 2023)

South Africa

Although South Africa does not have an explicit ESG disclosure regime, listed companies on the Johannesburg Stock Exchange (JSE) are required to disclose an integrated report since 2010, as recommended by the King IV Report (Institute of Directors Southern Africa, 2016). Compliance with the King Code is part of the JSE Listing Requirements (JSE, 2022) and includes disclosure requirements on financially material ESG information in annual reports on a comply-or-explain basis, as well as mandatory external assurance (Ioannou and Serafeim, 2017, p. 9; Stolowy and Paugam, 2018, p. 527). In addition, mineral companies are mandated by the Listing Requirements to include information regarding their environmental management and funding (Paragraph 12.13(iii)(13)), as well as major environmental issues (Paragraph 12.10(h)(viii)) in their annual reports (JSE, 2022).

2.3. Comparative analysis of mandatory ESG disclosure regulatory approaches

ESG regulatory regimes vary widely across countries, as shown in Section 2.2. The EU has adopted the most comprehensive mandatory ESG disclosure regime. The CSRD will expand the scope of companies covered and the range of issues to be reported to an unprecedented degree (European Parliament, 2023). Although the US does not yet have a wide-ranging federal ESG disclosure regime, the SEC has issued a proposal for comprehensive climate-related disclosure requirements in 2022 (SEC, 2022b). Progress is also being made at the state level, with California in particular moving forward with its own regulatory framework (California Government, 2023a, 2023b). In APAC, the level of mandatory ESG requirements varies. In China, only a subset of listed companies and state-owned enterprises are covered by comprehensive ESG disclosure requirements (SASAC, 2008; SSE, 2008; SZSE, 2008). In contrast, India goes beyond mere disclosure requirements by requiring certain companies to actively invest in ESG-related activities (MCA, 2013). While Hong Kong, Japan, and Malaysia have implemented more comprehensive reporting regimes (Bursa Malaysia, 2022; FSA, 2023; HKEX, 2019), Australia is just starting to move from voluntary to mandatory ESG reporting (Australian Government, 2023). Latin America is lagging behind, but countries such as Chile and Brazil are increasingly recognizing the importance of ESG transparency and implementing ESG disclosure requirements (CMF, 2021; CVM, 2023).

The remainder of this section provides an overview of the differences and similarities between mandatory ESG disclosure regulations along several dimensions, namely regulator type, basis of reporting, reporting standards, scope of companies affected, content, materiality assessment, disclosure

format, level of enforcement, and institutional and cultural factors.²¹

Regulator type: First, ESG regulations can be issued by the government (or regulatory bodies such as securities regulators) or by a stock exchange. For example, the EU disclosure directives are transposed into national law by member states' governments, while the Hong Kong regulation is issued by the stock exchange. The type of issuer is likely to influence the effectiveness of the regulation. According to Leuz (2010, p. 12), government regulation is characterized by stricter enforcement mechanisms, as governments have the power to impose criminal penalties on companies that do not comply with the law. Non-compliance by companies subject to stock exchange regulations, on the other hand, can usually only be punished by fines or delisting. While exchanges can tailor regulations to the companies they affect, government regulations, which cover a broader range of companies, benefit from standardization and network effects. Ernstberger et al. (2021, p. 5) provide evidence that government-imposed ESG disclosure requirements lead to a greater reduction in information asymmetry than do stock exchange requirements. Similarly, Krueger et al. (2023, p. 4) find that liquidity improves significantly more for firms subject to government requirements than for firms subject to stock exchange requirements.

Basis of reporting: Disclosure requirements can be implemented on a comply-or-explain basis or on a fully mandatory basis. The former is used by many jurisdictions to give companies the flexibility to either comply with the requirements or explain why they are not complying. For example, the NFRD allows companies to provide reasons for not implementing certain ESG-related policies (EPRS, 2021, p. 3). Some countries, such as Hong Kong, also opt for a mixed approach, which includes both mandatory and comply-or-explain disclosure requirements (HKEX, 2019, p. 1). In contrast, neither the SEC's nor the Australian government's proposed climate disclosure rules currently include a comply-or-explain principle (SEC, 2022b).

On the one hand, comply-or-explain provisions reduce compliance cost as companies are exempt from reporting information that is not relevant to them. The principle may also lead to greater acceptance by affected companies compared to a stricter, mandatory disclosure regime. Finally, the explanations provided by companies in the event of non-compliance could make it easier for policymakers to identify inefficiencies and weaknesses in the regime, enabling them to review and improve the relevant areas of the law. (Ho, 2017, pp. 345-347)

On the other hand, comply-or-explain may reduce the number of companies that do comply with the regulation, as companies can simply choose to provide an explanation. This reduces the effectiveness of the mandate and leads to

²¹ It is important to note that I do not attempt to evaluate the regulatory regimes in terms of potential strengths and weaknesses, nor do I analyze the potential costs and benefits for the companies subject to the mandates.

an increase in the quantity, but not the quality of disclosures. (Ho, 2017, p. 331) According to Haji et al. (2023, p. 181) the impact of these different reporting models on regulatory outcomes remains unclear.

Reporting standards: Many organizations have developed ESG disclosure standards with the goal of increasing harmonization of ESG disclosures. However, this has resulted in a fragmented ESG standards landscape: Common standards have been developed by the GRI, the TCFD²², the Sustainability Accounting Standard Board (SASB), the Climate Disclosure Standards Board (CDSB), and the International Integrated Reporting Council (IIRC), among others. (H. B. Christensen et al., 2021, p. 1177) To improve the comparability and consistency of information disclosed across companies and jurisdictions, the International Financial Reporting Standards (IFRS) Foundation has established the ISSB to consolidate the variety of existing standards. The ISSB has published their first set of standards in June 2023. (IFRS, 2023a, 2023b) Multiple jurisdictions have already announced that they are considering to adopt the ISSB standards, including the UK, Australia, and Canada, which would make compliance mandatory (EY, 2023, p. 2). Other major reporting frameworks are emerging in the EU and in the US. With the entry into force of the CSRD, the EU will require reporting according to the new ESRS standards developed by EFRAG (European Commission, 2023c). In the US, the SEC has developed its own reporting standards with its Proposed Rule on Climate-Related Disclosure (SEC, 2022b). The ISSB, ESRS, and the SEC Proposed Rule are all aligned with the TCFD recommendations. Although differences in requirements remain (i.e., regarding the scope, materiality definition, assurance requirements, and content to be disclosed), there is a trend towards more comparable and consistent reporting. Especially companies operating in multiple jurisdictions would highly benefit from a harmonization of reporting requirements. (EY, 2023, pp. 1-2)

Scope of affected companies: ESG disclosure requirements differ in terms of the scope of companies covered. Criteria used to define the scope include size thresholds (e.g. number of employees, annual revenue), whether a company is listed or not, or industry membership. Regulations can either cover a broad range of companies (e.g. the NFRD) or be targeted at a specific group of companies, such as large emitters (e.g. the US GHGRP).

Scope of content to be disclosed: Broad ESG disclosure regimes, such as the NFRD, cover multiple ESG issues, while targeted regulations, such as Section 1502 of the Dodd-Frank Act, focus on specific ESG concerns. Broad ESG regimes also vary in scope. For example, the SEC's proposed rule and the ISSB standards focus primarily on climate-related disclosures, although the ISSB is expected to expand the scope to include other issues. The ESRS, on the other hand, are more comprehensive, covering other environmental aspects such

as water use and pollution, social aspects such as those related to employees and consumers, and governance aspects such as business conduct (EY, 2023, p. 4). In addition, some regulations provide considerable detail on the content to be provided (e.g., Hong Kong's ESG Reporting Guide), while other requirements are more general (e.g., Brazil's Resolution No. 59). While prescribing the specific ESG matters, metrics and KPIs to be disclosed allows for better comparability of information across companies, the latter approach reduces the regulatory burden on companies by giving them the flexibility to assess which information is most material to them (European Commission, 2021, pp. 13–14).

Materiality: Reporting requirements also differ in their underlying definition of materiality, which is the concept used to identify relevant ESG issues for disclosure. In general, ESG issues can be material from a financial (or outside-in) perspective and/or from an impact (or inside-out) perspective. The financial materiality perspective considers those ESG issues that (potentially) have a material impact on the financial performance of the company. The impact materiality perspective requires companies to disclose their impact on the environment and society. (European Commission, 2021, pp. 15–16)

Most jurisdictions, including the SEC's proposed rule and countries adopting the ISSB standards, require reporting under the financial materiality perspective. The EU is the first jurisdiction to adopt a double-materiality approach, which requires companies to report on issues that are material from both perspectives combined, as well as issues that are material only from a financial or only from an impact perspective. (EY, 2023, p. 15)

Disclosure format: ESG regulations differ in their requirements for the location of disclosure. Information can be disclosed either as a separately published ESG report, as part of the annual report or as part of an integrated report. The integrated report is a combination of the first two disclosure formats, in that both ESG-related and financial information are comprehensively covered. (Ernstberger et al., 2021, p. 7) Hence, the integrated report aims to provide a broader view of the material financial, environmental, and social impacts on the company's value creation process (Cohen & Simnett, 2015, p. 16). For example, South Africa requires ESG information in the form of an integrated report since 2010, while Malaysia requires ESG information to be included in annual reports. Companies subject to the NFRD can either include ESG information in their annual report or as a separate report. Ernstberger et al. (2021, p. 5) find that regulations that require disclosure of ESG information in dedicated, standalone reports result in a greater reduction in information asymmetry than mandates that require reporting as part of annual or integrated reports. They argue that investors can access the ESG information more easily when presented separately (and not included in predominantly financial reports). However, there is still limited empirical evidence on the effect of different types of disclosure format (Haji et al., 2023, p. 193).

²² The TCFD takes a slightly different approach, being more of a framework to provide guidance to others, rather than being a standard in itself (Murray, 2021)

Targeted regulations may also require information disclosure on the company's website, or direct reporting to a specific authority (European Commission, 2021, p. 19). For example, companies subject to the US GHGRP have to disclose GHG emissions to the EPA, which will verify and consolidate the information and make it available on their own website (US EPA, 2023a).

Level of enforcement: The level of enforcement is likely to have a large impact on the effectiveness of the respective ESG disclosure rules (H. B. Christensen et al., 2021, p. 1190). The imposition of assurance requirements or penalties for non-compliance are two ways in which regulators enforce disclosure requirements.

According to Ioannou and Serafeim (2017, p. 12), external assurance is likely to reduce intentional misreporting in the form of missing or incorrect information. However, most regulations currently do not mandate external verification of ESG disclosures (Haji et al., 2023, p. 181). While the NFRD only requires assurance on the provision of ESG information, but not on the specific content (European Parliament and Council, 2014, p. 3, Recital 16), the CSRD requires more comprehensive external assurance (European Commission, 2023d).²³ Companies subject to Section 1502 of the Dodd-Frank Act are required to obtain external verification when claiming not to use conflict minerals (SEC, 2017). The SEC's current proposed rule requires external verification of Scope 1 and Scope 2 emissions for certain large companies (SEC, 2022a, p. 3).

Regulators can also pose penalties on companies in case of non-compliance to enforce disclosure rules. However, many ESG disclosure regulations do not state any specific sanction in case of non-compliance by reporting companies (Ioannou & Serafeim, 2017, p. 3). The NFRD, for example, does not prescribe specific penalties but urges member states to ensure that an effective enforcement mechanism is installed (European Parliament and Council, 2014, p. 2, Recital 10). Under the CSRD, penalties will also be subject to specification by member states (European Parliament and Council, 2022, p. 38, Recital 73). Some stock exchanges, such as the SSE and the SZSE in China, have made ESG reporting a listing requirement, meaning that companies risk delisting if they do not comply (Y.-C. Chen et al., 2018, p. 187). According to Ioannou and Serafeim (2017, p. 11), however, no company has yet been delisted for failing to publish ESG information.

Institutional environment and cultural aspects: The institutional environment of the country, consisting of formal and informal institutions, also influences the level of effectiveness of a regulation. A country with a strong formal enforcement mechanism is characterized by an effective regulatory environment and government, as well as strict rules of law.²⁴ While countries such as Germany, the US, and Aus-

tralia have strong formal institutions, regulations in India, Argentina, and South Africa are less likely to be strictly enforced (Krueger et al., 2023, pp. 44–45). Informal institutions refer to the society's norms and values. In the ESG context, social and environmental values in particular can influence the extent to which companies comply with disclosure requirements. (Krueger et al., 2023, pp. 4–5)

For instance, Krueger et al. (2023, p. 5) find that ESG mandates have a stronger positive effect on stock liquidity in countries with strong informal institutions, while a strong formal environment does not seem to have an impact. Mittelbach-Hörmanseder et al. (2021, p. 311) provide evidence that a stronger regulatory framework and a higher level of enforcement in EU member states have a positive impact on the value relevance (i.e. the explanatory power) of disclosed ESG information under the NFRD.

Similar to the institutional environment, cultural differences are also likely to influence the effectiveness of ESG disclosure regulation. Mittelbach-Hörmanseder et al. (2021, p. 335) find that differences in national culture, as measured by Hofstede's (1983) cultural dimensions, affect the value relevance of ESG disclosure (except for environmental issues). Their results suggest that power of distance (i.e., the degree of acceptance of hierarchy and authority) and individualism (i.e., the extent to which individuals prioritize themselves over the group) have a significant positive impact. Conversely, masculinity (i.e. the degree of competition in a society) and uncertainty avoidance (i.e. the degree of tolerance for ambiguity and risk) have significantly negative effects.

Overall, it is important to consider these differences in regulatory frameworks and the specific institutional and cultural contexts of each jurisdiction when interpreting the findings on the impact of ESG disclosure requirements on covered companies. This is particularly important when assessing their relevance and applicability to other regulatory environments.

3. Impact of ESG disclosure regulation on affected companies

This chapter reviews the relevant literature on the impact of ESG disclosure regulation on affected companies. I limit the scope of the analysis in two ways. First, I analyze the effects of ESG disclosure rather than the effects of firms' underlying ESG performance (although there is a close link between disclosure and firms' activities) (H. B. Christensen et al., 2021, p. 1184). Second, I focus on mandatory requirements, consistent with my review of mandatory regulatory developments in Chapter 2. Studies focused on mandatory disclosure regimes mitigate the dual selection bias found in studies of voluntary disclosure. That is, companies can first

²³ Under the NFRD, EU member states can choose to voluntarily implement mandatory assurance requirements, as well as whether penalties will be imposed in case of non-compliance. (CSR Europe and GRI, 2017, p. 7)

²⁴ Government effectiveness is defined as the overall efficiency of the gov-

ernment in implementing policies and its credibility in committing to them. The rule of law refers to the quality of the country's legal system and its ability to enforce the law. (Krueger et al., 2023, p. 18)

voluntarily choose which ESG activities to engage in and second, companies then choose whether and what ESG information to disclose (H. B. Christensen et al., 2021, p. 1198).

My literature review covers three areas of impact. First, the impact on the ESG reporting practices of affected firms (see Section 3.1); second, the financial impact of ESG disclosure regulation (see Section 3.2); and third, the ESG-related behavior of affected firms, also referred to as real effects²⁵ in the literature (see Section 3.3). A summary of the studies reviewed is provided in Appendix 2. Results vary across institutional contexts, likely reflecting the differences in mandates and country characteristics.

3.1. Impact on ESG disclosure practices

Several studies examine whether mandatory ESG disclosure regulations achieve their primary goal of increasing ESG transparency, as measured by changes in the quantity and quality of ESG reporting. The majority of studies suggest that ESG disclosure mandates lead to an increase in the quantity of disclosure (e.g., Cuomo et al., 2022; Fiechter et al., 2022; Ioannou and Serafeim, 2017), although this is not unanimously supported (Cordazzo et al., 2020). ESG disclosure mandates appear to have a positive or no effect on ESG reporting quality (D. M. Christensen et al., 2022; Ottenstein et al., 2022).

3.1.1. Quantity of ESG disclosure

ESG disclosure requirements potentially increase the number of reporting companies by requiring disclosure from companies that have not previously engaged in voluntary ESG reporting. They may also increase the depth of disclosures by requiring more comprehensive information beyond what is voluntarily disclosed. (Gulenko, 2018, p. 8)

However, the precise impact on the level of disclosure is unclear a priori. Many regulations operate on a comply-or-explain basis, allowing firms to avoid disclosure by justifying non-compliance (e.g. due to unreasonable costs or complexity of disclosure). In addition, the lack of robust enforcement mechanisms and penalties for non-compliance, as well as vague guidelines for reporting metrics, further contribute to this ambiguity. (Haji et al., 2023, p. 184); Ioannou and Serafeim, 2017, p. 11

Despite the comply-or-explain principle, several studies suggest that the EU's NFRD has led to an increase in ESG disclosure. For example, Fiechter et al. (2022, pp. 1509–1512) provide evidence that both the number of companies publishing ESG reports and the breadth of the individual reports increase. These effects already materialize after the passage of the directive (before companies are required to comply) and are stronger for firms with prior low levels of voluntary ESG disclosure. Similarly, Ottenstein et al. (2022, p. 56) find that the number of reporting companies and the

quantity of disclosed information per company increases after the NFRD comes into effect (2017 to 2018). Since the large public companies subject to the NFRD already tend to have high pre-mandate reporting levels, the effect is however less pronounced than expected (Ottenstein et al., 2022, p. 75). Cuomo et al. (2022, pp. 1–2) also find that more companies disclose ESG information following the NFRD's passage. Companies with certain characteristics show stronger positive effects, such as smaller firm size, stronger legal systems in the country of headquarter, and a higher number of analysts following. Mion and Adai (2020, p. 72) support these results as they find that the number of published ESG reports by affected companies in Germany and Italy increases in the first year of the NFRD's effectiveness.

Studies on mandates in France and the UK provide supporting evidence. Chauvey et al. (2015, p. 800) observe a significant increase in disclosure quantity following the introduction of the NRE in France, as measured by the amount of space used for ESG-related information in disclosure reports. Similarly, Hummel and Rötzel (2019, pp. 229–232) suggest that firms affected by the UK's Companies Act Regulations 2013 significantly increase disclosure levels. In additional analyses, they find a weaker positive effect for firms with high reporting incentives, as these companies tend to have higher pre-regulation voluntary disclosure levels.

In an international study, Ioannou and Serafeim (2017, pp. 4–5) also show that disclosure levels of companies in China, Denmark, Malaysia, and South Africa increase significantly after the implementation of respective mandatory disclosure regulations, as measured by ESG disclosure scores. The results do not seem to be influenced by the level of prior voluntary disclosure. Similar to the NFRD, the regulations are based on a comply-or-explain principle (Ioannou & Serafeim, 2017, p. 11).

In contrast, there is some evidence that the quantity of ESG information disclosed is not affected or even decreases after the introduction of disclosure requirements. For example, Cordazzo et al. (2020, p. 3476) find no evidence that affected Italian companies significantly increase the level of disclosure in the first year after the NFRD takes effect. In particular, companies with a history of voluntary ESG reporting seem to limit their disclosure even to the extent of meeting the required minimum standards. Similarly, companies reporting for the first time only disclose the minimum required. Bell (2021, pp. 23–25) analyses the impact of the *Indian Companies Act 2013*, which requires affected companies to disclose ESG-related information and to spend on ESG-related activities. The results do not suggest an improvement in the level of ESG disclosure, as measured by the number of ESG-related words in annual reports. In addition, the study finds that companies with previously lower disclosure levels relative to their peers significantly reduce disclosure levels after the regulation. This suggests that, on average, the mandate reduces disclosure levels compared to a voluntary environment. However, Bell (2021, p. 65) notes that the lack of improvement in disclosure may be due to weak enforcement in India. Therefore, it is questionable whether these re-

²⁵ Consistently with Leuz and Wysocki (2016, p. 530), real effects are defined as companies adapting their behaviour as a consequence of disclosure mandates, for example by reducing their resource consumption.

sults are generalizable to other regulatory environments with stronger enforcement mechanisms.

3.1.2. Quality of ESG disclosure

Research assessing the quality of mandatory ESG reporting often measures key aspects such as credibility, comparability, and completeness of reporting. It is important to examine whether mandatory disclosure improves reporting quality, as previous literature has mostly found low quality in voluntary ESG disclosures. (Haji et al., 2023, p. 184)

Several studies suggest that the NFRD positively impacts reporting quality. For example, Fiechter et al. (2022, pp. 1509–1510) find that ESG reports improve both in terms of credibility (as measured by the level of external assurance on disclosure) and comparability (as measured by the adoption of international reporting standards). These effects are more significant for companies with previously lower levels of voluntary reporting. Furthermore, the improvements already occur after the NFRD's passage in 2014 but before coming into effect in 2018. Ottenstein et al. (2022, p. 56) also find an overall improvement in reporting quality, which is driven by an increase in the credibility of ESG reports (as measured by external assurance levels). However, they find no increase in the adoption of international reporting standards, which suggests that the comparability of reports is likely to remain low. Mion and Adai (2020, pp. 72–74) find that reporting quality increases in the first year of the NFRD's effectiveness for German and Italian companies. In line with Fiechter et al. (2022) and contrary to Ottenstein et al. (2022), they find that comparability of ESG reports increases as more companies adopt international reporting standards (although the degree of compliance with reporting guidelines varies across companies). However, the results of the study should be considered indicative due to the limited scope of the dimensions included in the content analysis, the short time period covered, the small sample size and the lack of control for firm-level characteristics.

Examining different mandates across jurisdictions, Ioannou and Serafeim (2017, pp. 4-5) report an improvement in the quality of ESG disclosure by affected companies. This improvement relates to both credibility and comparability, as measured by changes in voluntary external assurance of ESG information and adoption of international reporting standards. D. M. Christensen et al. (2022, p. 160) also conduct an international analysis and find that companies subject to regulation significantly improve their reporting quality, as measured by the Bloomberg ESG disclosure score.

In contrast, other studies do not provide evidence for a positive impact of mandatory ESG disclosure on reporting quality. For example, Lock and Seele (2016, pp. 187–189) compare the quality of ESG reports of companies in European countries with mandatory non-financial disclosure regulations to those without such regulation (before passage of the NFRD). They measure the credibility of reports as defined by four characteristics: understandability, truth, sincerity, and appropriateness. The results do not provide evidence that disclosures from countries with ESG reporting

regimes are significantly more credible than those from countries without respective laws (Lock & Seele, 2016, p. 192). The study differs from those mentioned above in that it does not compare changes in disclosure quality before and after the passage of ESG disclosure mandates, but rather compares disclosure quality of companies in countries with regulation to those in countries without regulation. Therefore, it does not rule out the possibility that other country-specific factors influence the outcome.

Chauvey et al. (2015, pp. 800-801) study the change in quality of ESG reports in the years following the introduction of mandatory ESG disclosure in France in 2001 as mandated by the NRE. Although they find a slight improvement, disclosure quality remains at a low level. Their analysis further indicates that the number of companies reporting about negative ESG incidents decreases. However, the significance of the results is limited by the small sample size and by the fact that only two reporting years are included in the analysis.

Birkey et al. (2018, pp. 827–828) find that while companies affected by the CTSCA (the Californian disclosure mandate to address modern slavery in the supply chain) seem to generally comply with the disclosure mandate, reporting is rather symbolic. Companies do not provide detailed and substantial information with respect to the required disclosure areas.

3.1.3. Discussion of findings on corporate reporting practices

Overall, the majority of studies suggest that ESG disclosure regulation positively impacts disclosure quantity, both in terms of the number of companies reporting and the level of reporting per company (Fiechter et al., 2022; Ottenstein et al., 2022). Nonetheless, some studies do not provide supporting evidence (Cordazzo et al., 2020). In terms of disclosure quality, several studies find evidence for a positive impact of ESG disclosure regulation, as measured by higher credibility through voluntary external assurance (Fiechter et al., 2022; Ioannou & Serafeim, 2017; Ottenstein et al., 2022), higher comparability through the adoption of reporting standards (Fiechter et al., 2022; Ioannou & Serafeim, 2017; Mion & Adai, 2020), or higher overall ESG disclosure scores (D. M. Christensen et al., 2022).

However, other studies do not find evidence that ESG disclosure requirements increase the credibility (Lock & Seele, 2016) or comparability (Ottenstein et al., 2022) of reports. According to Chauvey et al. (2015, p. 800), firms seem to be selective in their disclosures, as the number of firms reporting about negative ESG incidents decreases post-regulation. In addition, Birkey et al. (2018, p. 828) finds that disclosure is rather symbolic. Despite the mixed evidence regarding reporting quality, broad ESG disclosure regimes like the NFRD seem to have a positive impact (Cuomo et al., 2022; Fiechter et al., 2022).

The impact of regulation on disclosure quantity and quality seems to be more pronounced for companies that are more affected by the regulation. That is, firms with certain firm-level characteristics such as lower prior voluntary disclosure

levels²⁶ (Fiechter et al., 2022, p. 1510; Hummel and Rötzel, 2019, p. 229), smaller firm size (Cuomo et al., 2022, p. 2; Ottenstein et al., 2022, p. 75), or firms located in countries with stronger legal systems (Cuomo et al., 2022, p. 2). Fiechter et al. (2022, p. 1510) also suggests that companies improve their reporting practices in anticipation of the new regulations taking effect.

3.2. Financial impact of ESG disclosure regulation

When ESG disclosures reveal new and relevant information, they are expected to provide capital market benefits such as higher firm valuation, increased liquidity, and lower cost of capital. However, ESG disclosure is also associated with costs and risks, including the potential revelation of strategic business details. In addition, greater ESG transparency may increase legal or litigation risks. As a result, the ultimate impact of ESG disclosure on a company's financial market position and profitability is not readily predictable. (H. B. Christensen et al., 2021, p. 1208)

The first three parts of this section focus on the capital market effects of ESG disclosure requirements, including on firm value (Section 3.2.1), stock liquidity and information asymmetry (Section 3.2.2), and the cost of capital (Section 3.2.3). Section 3.2.4 focuses on the impact on firms' financial operating performance, as reflected in changes in firm profitability. Section 3.2.5 discusses the findings on the financial impact of ESG regulation.

3.2.1. Firm value

ESG disclosure can increase firm value if it provides useful information to investors, as discussed above. However, mandatory disclosure could also lead to companies being forced to provide less beneficial information, potentially harming firm value (Mittelbach-Hörmanseder et al., 2021, p. 316). Market perceptions of high compliance costs, including those for preparing ESG reports and adopting ESG-related governance practices, could also reduce firm value (Ioannou & Serafeim, 2017, p. 5). However, negative market reactions may be mitigated if the market expects a symbolic response to ESG regulation. In addition, costs may decline over time as firms become more efficient in complying with the requirements. (Haji et al., 2023, pp. 185–186)

Researchers have used a variety of approaches to study the impact of ESG disclosure requirements on firm value. One strand examines the longer-term valuation effects on companies following the introduction of regulation. Another examines capital market reactions to events that increase or decrease the likelihood of adoption of ESG disclosure mandates. The third examines the impact on firm value immediately after mandatory reporting begins.

Long-term valuation effects: With respect to the first strand of research, studies in different institutional settings

show mixed results. Ioannou and Serafeim (2017, pp. 5-6) find that increased transparency following ESG mandates is significantly positively associated with higher firm value (as measured by Tobin's Q) in different regulatory environments, including China and South Africa. This effect is attributed to both environmental and social disclosures. Krueger (2015, pp. 2–4) finds that the UK's 2013 GHG emissions disclosure mandate has a positive effect on firm value (as measured by Tobin's Q) of companies that did not voluntarily disclose emissions before the mandate. The positive effect is stronger for large firms and those in carbon-intensive industries. This suggests that investors particularly value the transparency of companies with high environmental impacts. The impact is already observed in 2011, before the new requirements take effect (Krueger, 2015, p. 19). Swift et al. (2019, pp. 424–426) report that firms with greater transparency experience improved stock market valuation (as measured by Tobin's Q) following the conflict minerals disclosure rule in Section 1502 of the Dodd-Frank Act, with effects also occurring before compliance is required.

Contrasting evidence finds a negative impact on firm value. For example, Lu et al. (2021, p. 1495) and Y.-C. Chen et al. (2018, pp. 176–177) show a significant decrease in firm value of Chinese firms subject to the 2008 ESG disclosure rules of the Shenzhen and Shanghai stock exchanges (as measured by price-to-book ratio and Tobin's Q, respectively). Mittelbach-Hörmanseder et al. (2021, p. 324) find that the introduction of the NFRD turns the previously positive or neutral relationship between ESG disclosure and firm value into a significantly negative one (as measured by share price). This decline is attributed to the involuntary disclosure of less beneficial information, investor anticipation of additional costs, content ambiguity, and lack of external assurance (Mittelbach-Hörmanseder et al., 2021, p. 327).

The institutional context also influences this relationship, with ESG disclosure showing lower value relevance in countries with high ESG awareness and labour protection, but higher relevance in countries with stronger legal systems and enforcement mechanisms (Mittelbach-Hörmanseder et al., 2021, pp. 335-336).

Valuation effects around relevant regulatory events: Most studies find that markets react negatively to key events that increase the likelihood of mandatory ESG reporting adoption. Grewal et al. (2019, p. 3062) observe a significant decline in the value of affected firms around three important events leading up to the passage of the NFRD in the EU, as measured by cumulative abnormal returns (CAR)²⁷. Firms with higher pre-mandate ESG disclosure and performance show a less pronounced negative reaction. Companies with above-median ESG disclosure and performance even experience positive reactions, suggesting that investors are mostly

²⁶ With the exception of Ioannou and Serafeim (2017, p. 11), who find no difference in disclosure levels between high and low reporting companies prior to regulation.

²⁷ The three events are considered to increase the likelihood of the introduction of mandatory ESG disclosure in the EU, namely (1) the European Commission presenting the proposal, (2) the European Council agreeing on the proposal, and (3) the Commission adopting the proposal. (Grewal et al., 2019, p. 3067)

concerned about higher costs for firms with lower ESG performance. Y.-C. Chen et al. (2018, pp. 184–186) find more negative market reactions to the announcement of mandatory ESG regulation in China for regulated firms (as measured by CAR). Unregulated firms also experience a negative (but less pronounced) market reaction, suggesting that investors expect regulation to be extended to currently unaffected firms.²⁸ According to Manchiraju and Rajgopal (2017, pp. 1260–1261), affected firms experience a negative stock market reaction around several events related to the passage of India's ESG disclosure and spending mandate (as measured by CAR and stock prices). The negative impact is mitigated for firms with higher advertising spending, which is a proxy for greater consumer awareness. Birkey et al. (2018, p. 834) report significant negative market reactions around major legislative events leading to the passage of California's 2010 CTSCA, especially for larger firms and those with higher supply chain risks (as measured by CAR).

Event studies related to the enactment of the US Dodd-Frank Act also observe negative market reactions around events increasing the likelihood of passage. Healy and Serafeim (2020, pp. 120–121) find stock price declines for companies affected by Section 1504 of the Act, which mandates disclosure of government payments in the extractive industries. The negative effects are stronger for firms with previously low government payment transparency, firms with higher foreign assets, firms that lack sufficient anti-corruption measures, and firms that operate in countries with high expropriation risk or that prohibit government payment disclosure. However, a small sample size of 26 firms limits the transferability of the results.

Similarly, Elayan et al. (2021, pp. 22–25) observe negative cumulated average abnormal returns (CAAR) for firms using conflict minerals around various announcements that are thought to increase the likelihood of passage of Section 1502 of the Act. The negative impact is stronger for less transparent companies, high-tech companies, companies dependent on conflict minerals in their production process, and companies with a history of human rights issues. However, companies with strong reputations and good governance experience less of a negative impact, as investors perceive them as less likely to be involved in human rights abuses.

In contrast, Cousins et al. (2020, pp. 32–34) find on average no effect on stock returns of affected companies around events leading to the passage of the UK Modern Slavery Act 2015. Firms with higher slavery risk experience more negative effects, while firms with better labor protection standards experience more positive stock price responses.

H. B. Christensen et al. (2021, p. 1200) raise issues with the event study approach. Market reactions may reflect anticipation not only of the effects of ESG disclosure, but also of actual changes in firms' ESG activities. In addition, it is difficult to identify specific dates when investor expectations

change. Finally, they point out the challenge of disentangling the effects of simultaneous external shocks, all of which can reduce the accuracy of the results.

First-time ESG reporting: Research on market reactions to first-time mandatory ESG reporting shows mixed results. Jouvenot and Krueger (2019, p. 7) find that among companies subject to the UK's GHG emissions disclosure mandate, those with lower emissions benefit from a positive market reaction (as measured by abnormal stock returns), while high-emission companies experience negative abnormal returns. Elayan et al. (2021, pp. 34–35) find that negative market reactions to initial disclosure under Section 1502 of the Dodd-Frank Act on conflict minerals disclosure are limited to firms with unclear disclosures, those with a history of human rights abuses, and those sourcing from the DRC or neighboring countries. Conversely, companies with risk mitigation strategies experience more positive reactions.

Cordazzo et al. (2020, pp. 3480–3481) find no evidence for a negative impact of first-time ESG disclosure on share prices of Italian companies subject to the NFRD. This may be due to unchanged disclosure levels of these companies after regulation (see Section 3.1.1). Veltri et al. (2020, pp. 2227–2230) analyze the impact of a subtopic of ESG reporting, namely ESG-related risk disclosure, on market value in the same setting and find a positive effect, although their findings are limited by a small sample size of 51 firms. While Veltri et al. (2020) analyze the impact of differences in risk disclosure levels post-mandate, Cordazzo et al. (2020) analyze the value relevance of ESG information pre- and post-regulation.

3.2.2. Stock liquidity and information asymmetry

In theory, ESG disclosure requirements increase transparency and reduce information asymmetry, which may positively impact stock liquidity.²⁹ However, these effects may not materialize if the disclosed ESG information lacks validity, comprehensibility, or financial relevance. (Krueger et al., 2023, p. 2)

Multiple studies observe improvements in stock liquidity and decreases in information asymmetry following mandated ESG disclosure. Krueger et al. (2023, pp. 35–36) provide evidence that several ESG disclosure requirements worldwide lead to positive stock liquidity effects. These are stronger for regulations enacted by governments (as opposed to stock exchanges) and regulations that do not adopt a comply-or-explain principle. While strong informal institutions (i.e., societal norms and values) also lead to a more pronounced positive effect on stock liquidity, no such effect was observed for jurisdictions with strong formal institutions (i.e., regulatory environment and government). Companies that did not disclose ESG information on a voluntary basis pre-mandate have the strongest increase in stock liquidity. These findings are in line with the results from Barth et al. (2017, p. 53).³⁰

²⁸ The study design differs slightly from other event studies in that it considers only the announcement of the regulation, as opposed to several events that increase the likelihood of its passage.

²⁹ Stock liquidity and information asymmetry are inversely related, meaning that a lower information asymmetry is often associated with a higher stock liquidity, and vice versa (Barth et al., 2017, p. 44).

³⁰ In contrast to most studies focusing on mandatory ESG disclosure

They provide evidence that stock liquidity increases for affected companies with higher quality disclosures following the mandatory integrated reporting disclosure regulation in South Africa in 2010, while information asymmetry (measured in terms of bid-ask spread) decreases.

In an international study setting, Ernstberger et al. (2021, pp. 19–21) find that companies subject to different ESG reporting regimes have a 18% stronger decrease in information asymmetry (as measured by bid-ask spread) compared to unaffected companies. In line with Krueger et al. (2023), they report stronger effects for regulations implemented by governments, and no significant effects for stock exchange regulations. The same applies to regulations requiring standalone ESG reports, whereas disclosure as part of annual or integrated reports do not have significant effects. In addition, the effects are stronger for firms in industries with a high proportion of affected companies, which increases the comparability of reports for investors. Ernstberger et al. (2021, pp. 24–26) also find stronger reductions in information asymmetry for companies using specific reporting guidelines (e.g., the GRI).

3.2.3. Cost of capital

If disclosure requirements reduce the information asymmetry between the capital market and the companies concerned, investor uncertainty would be reduced, which may lead to a reduction in the cost of capital (H. B. Christensen et al., 2021, p. 1202). In a voluntary setting, many studies provide corroborating evidence of this effect (e.g., Dhaliwal et al., 2011; Eliwa et al., 2021; Raimo et al., 2021).

Although there is limited research on mandatory settings, prior studies provide evidence in support of this notion. For example, Cuomo et al. (2022, pp. 12–13) find a reduction in cost of equity and systematic risk for affected companies after the introduction of the NFRD. Similarly, Xu et al. (2021, p. 2203) show that companies subject to the SSE and SZSE ESG disclosure requirements in China have a lower cost of debt, relative to unregulated companies. These effects are stronger for companies with longer ESG reports and companies that publish higher quality ESG reports. In contrast, Barth et al. (2017, p. 54) find no significant effect on cost of capital of affected companies in their study regarding the introduction of mandatory integrated reporting regulation in South Africa for JSE-listed companies.

H. B. Christensen et al. (2021, pp. 1211-1212) argue that lower cost of capital through disclosure mandates can also have real effects. First, a lower cost of capital can incentivize firms to invest, as it typically sets the threshold for new investments. Second, increased ESG transparency allows investors to assess a company's ESG risk exposure, which has an influence on the companies' cost of capital. This can prompt

reporting companies to mitigate risk, align with investor preferences, or generally improve ESG performance.

3.2.4. Firm profitability

Empirical evidence suggests a correlation between higher ESG engagement and improved operational performance, including higher productivity and sales (e.g., S. Chen et al., 2023, p. 9; Flammer, 2015, p. 2550). If mandatory disclosure encourages companies to adopt more sustainable practices (see Section 3.3), this implies a potential positive impact of ESG disclosure on financial performance. On the other hand, ESG disclosure is also associated with higher costs (Ioannou & Serafeim, 2017, p. 5). Based on the rationale discussed in Section 3.2.1 on firm value, companies would have already engaged in voluntary ESG disclosure and activities if they were beneficial to their financial performance, suggesting that a mandate would be detrimental to their performance (e.g., Y.-C. Chen et al., 2018, p. 172). This may also explain the mostly positive effect on financial performance found in studies based on a voluntary regulatory context, again pointing to a potential selection bias of these studies (H. B. Christensen et al., 2021, p. 1198). The evidence on the impact of ESG disclosure requirements on profitability (as measured by accounting metrics such as return on assets (ROA), return on equity (ROE), gross margin, or sales) is inconclusive.

Several studies suggest that ESG disclosure mandates negatively impact profitability of affected firms. For example, Y.-C. Chen et al. (2018, pp. 175–177) find a significant decline in the profitability of firms affected by the 2008 ESG disclosure regulations of the Chinese SSE and SZSE. Specifically, they report a 26% (20%) decline in ROA (ROE) compared to non-affected companies. Revenues and CapEx also decline after the mandate, while OpEx and impairment charges increase. The authors attribute these effects to potential plant closures or increased spending on pollution control and personnel following the mandates. Similarly, Lu et al. (2021, p. 1495) report a decline in corporate profitability that is noticeable two years after the regulation. The decline in ROA and ROE affects both state-owned enterprises (SOEs) and non-state-owned enterprises (NSOEs). This finding contrasts with Y.-C. Chen et al. (2018, pp. 182–183), who find the decline in profitability mainly in SOEs.³¹ Lu et al. (2021, p. 1510) suggest that the difference in findings may be due to the shorter time frame of the study of Y.-C. Chen et al. (2018), which may not have captured the longer-term effects seen in their own research.

In the US, H. B. Christensen et al. (2017, pp. 3-5) find that the requirement for SEC-registered companies to include mine safety information in financial statements under Section 1503 of the Dodd-Frank Act leads to a decrease in labor

regimes, the study of Barth et al. (2017, p. 64) is slightly different in that it focuses on a regulation mandating the adoption of integrated reporting. An integrated report comprises both financial and sustainability-related information (Barth et al., 2017, p. 45).

³¹ Agency problems are likely to arise in state-owned enterprises between the government and minority shareholders as well as between the SOE managers and shareholders, as explained by Y.-C. Chen et al. (2018, p. 182).

productivity. The decrease in productivity results in an increase in labor costs of approximately 0.9% of total revenues. Because the safety information was already published online by the government before the mandate, the study measures the incremental effect of including the information in financial reports, i.e., the effect of increased public awareness.

In contrast, Downar et al. (2021, p. 1166) find that requiring UK firms to disclose GHG emissions does not lead to a significant change in operating performance (as measured by gross margin). In particular, neither costs nor sales increase significantly for companies subject to the Companies Act 2006 Regulations 2013. Therefore, they do not support the notion that ESG disclosure regulation has a negative impact on the operating performance of affected companies.

Swift et al. (2019, pp. 421–424) find that Section 1502 of the 2010 Dodd-Frank Act regarding the disclosure of conflict minerals positively affects the profitability of reporting companies. Specifically, firms with higher supply chain transparency experience a larger increase in ROA compared to firms with lower transparency. This improvement is driven by both better sales performance and reductions in direct and overhead costs. In addition, companies appear to be improving their operations even before the mandate takes effect, as they gain greater visibility into their supply chain during the process of preparing for the new disclosure requirements.

3.2.5. Discussion of findings on the financial impact

Overall, literature on the financial impact of ESG disclosure regulation provides mixed evidence. Evidence on the longer-term valuation effects of ESG reporting mandates (e.g., Ioannou and Serafeim, 2017; Lu et al., 2021; Mancharaju and Rajgopal, 2017; Swift et al., 2019), as well as on the first reporting year post-regulation (e.g., Cordazzo et al., 2020; Elayan et al., 2021; Jouvenot and Krueger, 2019; Velti et al., 2020) is inconclusive. Events increasing the likelihood of adoption of regulation are mostly found to negatively impact firm value, suggesting that the capital market expects the regulation to impose additional costs on reporting companies (e.g., Birkey et al., 2018; Y.-C. Chen et al., 2018; Grewal et al., 2019; Healy and Serafeim, 2020). However, mandatory ESG disclosure regulations seems to have rather positive impacts on stock liquidity and information asymmetry (Barth et al., 2017; Ernstberger et al., 2021; Krueger et al., 2023), as well as on the cost of capital of affected companies (Cuomo et al., 2022; Xu et al., 2021). In addition, there are mixed findings on the direction of impact of ESG disclosure regulations on companies' financial performance (e.g., Y.-C. Chen et al., 2018; H. B. Christensen et al., 2017; Downar et al., 2021).

Several studies report more pronounced effects for firms that are more affected by the regulation. Grewal et al. (2019, p. 3062), Elayan et al. (2021, p. 25) and Healy and Serafeim (2020, p. 121) find that the negative stock market reaction around key legislative events prior to ESG mandate adoption are more pronounced for firms with prior lower levels of disclosure. In contrast, Krueger (2015, p. 4) provides evidence

that companies without prior voluntary disclosure experience a more positive impact on longer-term firm valuation.

In addition, several studies suggest that companies that are expected to be more negatively affected by a disclosure mandate experience more negative valuation effects. This includes companies with higher emissions (Jouvenot & Krueger, 2019, p. 7) or higher supply chain risk (Birkey et al., 2018, p. 834), as well as companies with a history of social incidents (Cousins et al., 2020, p. 33) or human rights violations (Elayan et al., 2021, p. 34), or companies operating in high expropriation risk countries (Healy & Serafeim, 2020, p. 121). Companies with positive firm characteristics, such as good reputation, better ESG performance, or better labor protection standards, seem to be able to mitigate any potential negative effects from ESG disclosure regulation (Cousins et al., 2020, p. 33; Elayan et al., 2021, p. 22; Grewal et al., 2019, p. 3062).

Moreover, there is evidence that positive effects on stock liquidity and information asymmetry are stronger for government regulations (as opposed to stock exchange regulations), and for regulations without comply-or-explain provisions and those requiring disclosure in stand-alone reports (Ernstberger et al., 2021, pp. 19–21; Krueger et al., 2023, pp. 35–36). Finally, the strength of the institutional environment also seems to influence the outcome of ESG disclosure regulations (Mittelbach-Hörmanseder et al., 2021, p. 335).

Finally, Krueger (2015, p. 19) and Swift et al. (2019, p. 425) report that Tobin's Q of affected companies already increases before mandatory compliance with the regulation, which suggest anticipatory effects. In contrast, Lu et al. (2021, p. 1495) reports that the impact on profitability may be delayed and unfold over a longer period of time.

3.3. Impact on ESG-related company performance

Besides its primary goal to increase transparency for various stakeholders, ESG disclosure requirements also aim to foster sustainable corporate behavior by holding companies more accountable (Haji et al., 2023, p. 178). For example, the EU states that ESG disclosure plays a central role in creating a sustainable economy which protects the environment and society (European Parliament and Council, 2014, Recital 3), demonstrating the intention to encourage more responsible firm behavior (Fiechter et al., 2022, p. 1500).

According to H. B. Christensen et al. (2021, pp. 1212–1213), several mechanisms may drive the desired impact of ESG mandates on companies' ESG performance. First, ESG disclosure requirements make it easier for investors to monitor companies' ESG activities. As a result, investors can exert pressure if the company's activities are not aligned with their preferences, e.g. through shareholder activism or the sale of shares. Second, disclosing companies may be pressured by various stakeholders other than investors. The anticipation of potential negative reactions, such as reduced product demand or public shaming from stakeholders like consumers and social activists, may encourage companies to improve their ESG performance. Third, peer benchmarking could lead

to firms improving their ESG performance. Higher transparency of ESG practices enables firms to learn from their peers and encourages them to outperform competitors. Finally, ESG disclosure regulation could lead to the adoption of more ESG activities if this enhances firms' financial performance. The notion is that higher transparency leads to increased customer awareness of a firm's ESG performance. Customers may reward good ESG performance with higher trust and loyalty, which could positively impact financial performance.³²

On the other hand, legitimacy theory suggests that instead of improving ESG performance through actual adaptation of corporate operations, companies subject to ESG disclosure regulation may respond by strategically increasing disclosure in areas where they already have good ESG performance, and neglecting areas of poor performance in their reports (Haji et al., 2023, p. 188). Nonetheless, Bauckloh et al. (2023, p. 143) argue that while this is a common issue in voluntary reporting settings, the switch to a mandatory regulatory setting is likely to mitigate such legitimization strategies, and incentivize companies to implement real changes. Relevant studies of various jurisdictions provide mixed evidence on whether ESG disclosure regulation leads to the desired outcome of better ESG performance.

This section is organized into a review of studies that examine the impact on overall ESG performance (Section 3.3.1) and studies that examine the impact on specific outcomes related to the environmental (Section 3.3.2), social (Section 3.3.3) or governance pillar (Section 3.3.4). Section 3.3.5 discusses and summarizes the findings of the literature.

3.3.1. Overall ESG performance

According to Fiechter et al. (2022, pp. 1512–1514), companies subject to the NFRD increase their ESG activities in 2018, i.e. the first mandatory reporting year. This positive effect is on average only significant for the social dimension, and the effects are stronger for companies with previously lower levels of ESG-related activities and disclosure. Fiechter et al. (2022, pp. 1515–1517) argue that stakeholder pressure and the possibility to benchmark ESG performance against peers are potential underlying mechanisms for improved ESG performance. In addition, they observe that companies already begin to improve their ESG performance in 2014, i.e. after the passage of the NFRD but before the disclosure requirements become effective. The early onset of the effect could be due to three possible reasons. First, companies may reveal and potentially mitigate inefficiencies in their ESG policies whilst preparing for the new disclosure requirements. Second, increasing ESG awareness of the public could pressure companies to adopt more ESG-friendly behavior. Third, companies could increase ESG performance in anticipation of future stakeholder reactions, e.g. to maintain a high reputation with customers.

³² However, empirical studies does not provide unambiguous evidence on a positive impact on firm profitability (e.g., Y.-C. Chen et al., 2018; Lu et al., 2021), as discussed in Section 3.2.4

Cuomo et al. (2022, pp. 6–9) also find that the NFRD leads to a significant increase in ESG scores of affected companies. Contrary to Fiechter et al. (2022), their results suggest a positive impact on both social and environmental activities. Smaller firms show stronger improvements in ESG performance. As smaller firms tend to engage less in ESG-related activities prior to the NFRD (due to lack of resources), this aligns with the findings of Fiechter et al. (2022). Moreover, they find no evidence that requiring external assurance of ESG reports, as voluntarily adopted by some EU countries under the NFRD, significantly affects ESG performance. This finding is particularly interesting in light of the new CSRD, which includes external assurance requirements. As the improvements are observed in the period from 2015 to 2018, the results can also be interpreted as anticipation effects.

Cicchello et al. (2023, pp. 1124–1125) provide supporting evidence, as companies subject to the NFRD significantly increase in ESG scores in the period from 2018 to 2020. Thus, they measure the impact of the NFRD coming into force (as opposed to its passage).

In a different setting, Dharmapala and Khanna (2018, p. 104) find that firms subject to Section 135 of India's Companies Act of 2013 significantly increase their ESG activities (as measured by ESG spending). However, the Indian mandate is not a pure disclosure regulation, but also requires ESG-related spending. Interestingly, they find that out of the 100 largest firms in their sample, those companies already spending more than 2% on ESG prior to the mandate decrease their ESG-related expenditure following the introduction of the Act.

In their analysis of the evolution of Swedish companies' ESG performance between 2009 and 2018, Arvidsson and Dumay (2022, p. 1106) find no evidence of a significant improvement from 2015 onwards (i.e. after the NFRD's passage). They argue that their finding could be due to the fact that Swedish companies already had a comparatively high level of voluntary disclosure prior to the NFRD, which may have mitigated the impact of the directive. However, the sample includes only 27 listed companies (Arvidsson & Dumay, 2022, p. 1096).

3.3.2. Environmental performance

Studies focusing on environmental performance report a reduction in industrial wastewater pollution (Y.-C. Chen et al., 2018), GHG emissions (Downar et al., 2021; Jouvenot & Krueger, 2019; Tomar, 2023) and other types of pollution (Doshi et al., 2013; Gramlich & Huang, 2017).

Y.-C. Chen et al. (2018) and Gramlich and Huang (2017) provide evidence that the 2008 ESG disclosure regulations by Chinese SSE and SZSE have been effective in nudging firms to reduce their environmental impact. Y.-C. Chen et al. (2018, p. 170) analyze the impact on city-level pollution over the period 2006 to 2011. Their results show more significant reductions in wastewater and toxic sulfur dioxide emissions in cities with many companies affected by the regulation. Measuring changes in city-level pollution (as opposed

to company-level pollution) allows to observe the overall reduction in pollution, as it also considers a potential effect on pollution levels of unregulated companies in the same city. Gramlich and Huang (2017, pp. 4–5) examine the impact on company-level pollution between 2005 and 2013 and find a greater reduction for affected companies relative to unregulated companies. They observe the biggest reductions in water extraction, natural resource depletion and chemical emissions.

Doshi et al. (2013, pp. 1226–1228) examine whether affected firms reduce their environmental emissions following the expansion of the US EPA's TRI disclosure regulation, and find heterogeneous results. They find that the effectiveness of the disclosure mandate depends on the extent of internal and external pressure faced by the firm, as well as access to the necessary skills.³³ They also find that in sparsely populated regions, large firms reduce their emissions less than small firms, but no difference in performance is observed in dense areas. The authors argue that large firms have greater political power in sparsely populated regions, allowing them to resist institutional pressures to improve environmental performance. Finally, they find that subsidiaries of privately owned companies have greater emissions reductions than publicly traded companies. One potential reason may be that private companies are more willing to invest because they are not constrained by adverse capital market reactions.

Several studies investigate the impact of carbon disclosure mandates on GHG emission levels of affected companies. Downar et al. (2021) and Jouvenot and Krueger (2019) analyze the impact of the carbon disclosure requirement of the UK's Companies Act 2006 Regulations 2013, and find a reduction in GHG emissions for firms subject to the mandate. Downar et al. (2021, pp. 1139–1140) report absolute carbon reductions of 8%. Carbon intensity, as measured by absolute emissions relative to cost of goods sold and sales, improves by 13% and 10%, respectively. In addition, they find weaker reduction in absolute and relative emissions for companies with more complex company structures (Downar et al., 2021, p. 1166). In comparison, Jouvenot and Krueger (2019, pp. 4–5) report a reduction in absolute (relative) GHG emissions³⁴ of about 15% (17%). The variance in magnitude may be due to differences in the methodology: while Downar et al. (2021, p. 1141) analyze the change in emissions of individual facilities affected by regulation³⁵, Jouvenot and Krueger (2019, pp. 14–15) focus on affected firms' global emissions.³⁶ Both Downar et al. (2021, pp. 1170–1171)

and Jouvenot and Krueger (2019, p. 32) identify stakeholder pressure to drive emissions reductions, with the latter also arguing that managers learning from industry peers are likely to drive the reductions.

Importantly, many of the affected UK companies were already subject to the European Union Emission Trading System (EU ETS) before the UK's Companies Act emission disclosure requirement. As the EU ETS is a carbon pricing scheme, it is difficult to isolate the effect of the disclosure requirement from the effect of carbon pricing on emissions. (Bauckloh et al., 2023, p. 130; Downar et al., 2021, p. 1147) The EU ETS also requires companies to report emissions data from their individual installations to a public registry. Therefore, the observed effect is the incremental effect of aggregated disclosure as part of the annual statements, i.e. making GHG emission data more easily accessible to stakeholders. (Downar et al., 2021, p. 1139)

Evidence from the US also suggest that targeted carbon reporting regulation, like the 2010 US EPA's GHGRP has a positive impact on GHG emissions. Tomar (2023, p. 471) shows that absolute emissions decrease significantly by almost 8%, and carbon intensity by about 7%, which indicates that the reduction is not only due to a reduced economic activity. Tomar (2023, pp. 454–455) argues that reductions are driven by peer benchmarking and companies' anticipation of future GHG-related regulations. Only measuring emissions, without respective disclosure to the public, does also not lead to emission reductions. Bauckloh et al. (2023, p. 128) also analyze the impact of the GHGRP and provide evidence that affected firms significantly improve their carbon intensity ratio relative to unregulated firms. Contrary to Downar et al. (2021), Jouvenot and Krueger (2019), and Tomar (2023), they do not observe the same effect for absolute carbon emissions: affected firms have a weaker decrease in absolute emissions compared to unregulated firms. The study points to the need to further investigate why affected companies were unable to limit total emissions more than unaffected companies, despite better carbon intensity (Bauckloh et al., 2023, p. 144).

In contrast to the UK setting described above, most US companies did not disclose GHG data prior to the mandate. This means that the US studies observe the effect of initial disclosure rather than improved access to information. (Tomar, 2023, p. 452) Another difference is that emissions are published on the US EPA website, and do not have to be included in annual financial reports (Downar et al., 2021, p. 1141; US EPA, 2023a).

Matisoff (2013, pp. 588–589) analyses the impact of US state-level carbon disclosure mandates and finds no evidence of a reduction in absolute or relative GHG emissions. While these results contradict the studies mentioned above, it is important to note that they observe individual US state regulations.

Finally, the findings of Downar et al. (2021), Matisoff (2013), and Tomar (2023) are limited by the fact that they

from the Refinitiv ESG database.

³³ In particular, facilities located close to their headquarters or close to other facilities owned by the same parent company show a greater reduction in emissions. This supports the notion that these companies are subject to peer pressure and thus seek to maintain a good reputation. Companies that are close to sibling companies in the same industry also show greater improvements in environmental performance. This suggests that geographical proximity facilitates the transfer of skills. (Doshi et al., 2013, pp. 1226–1228)

³⁴ Jouvenot and Krueger (2019, pp. 3-4) define relative emissions as total emissions to tangible assets.

³⁵ Data as reported to and verified by the EU Transaction Log (EUTL).

³⁶ Data as aggregated by the non-profit organization CDP and estimations

examine the emissions of specific facilities that are subject to regulation, rather than the global emissions of affected companies. Therefore, they cannot rule out the possibility that affected companies simply shift emissions to facilities in unregulated regions. (Bauckloh et al., 2023, p. 130; Downar et al., 2021, p. 1139)

3.3.3. Social performance

Albeit limited, research focused on the social dimension suggests that ESG disclosure regulation has a positive impact on various outcomes, including employee-related matters in the US mining industry (e.g., H. B. Christensen et al., 2017) and other hazardous industries (e.g., Y.-C. Chen et al., 2018), and the due diligence of the supply chain (e.g., She, 2022).

In their analysis of 24 member countries of the Organization for Economic Cooperation and Development (OECD), Jackson et al. (2020, p. 334) provide evidence that companies subject to mandatory ESG regulation significantly increase activities related to the social pillar, with a stronger effect for companies with comparatively low levels of prior voluntary ESG activities. The analysis does not, however, provide evidence for a reduction in irresponsible actions.

H. B. Christensen et al. (2017, pp. 3-5) examine the real effects of Section 1503 of the 2010 Dodd-Frank Act, which mandates the disclosure of health and safety records in financial reports for SEC-registered mine-operating companies. They find that affected companies improve their compliance with safety regulations, as measured by a decrease in the number of citations issued to those companies for safety violations. In addition, they report a substantial decrease in mine injuries for companies affected by the mandate. As companies had already disclosed the required information online beforehand, the study measures the incremental effect of including the information in financial reports. Therefore, increased public awareness of the disclosed information is likely to create positive real effects.

Similarly, Y.-C. Chen et al. (2018, pp. 183–184) provide evidence that companies in dangerous industries experience a significant decline in workplace fatalities following the introduction of mandatory ESG disclosure regulation in China. Despite a relatively small sample due to limited data availability, these findings suggest that comprehensive ESG disclosure requirements lead to positive social externalities.

She (2022, pp. 400–401) finds an increase of supply chain due diligence for companies subject to California's CTSCA of 2010, which requires disclosure of due diligence processes regarding suppliers' labor abuses. The due diligence performance is proxied by the average human rights performance of the company's suppliers. The effects are stronger for companies facing external pressure from activist groups (e.g. Non-Governmental Organizations and shareholder advocacy groups) and for companies with suppliers located in countries considered to have low human rights standards. Also, the positive effects are more pronounced when the regulation enhances the comparability of the disclosed information.

3.3.4. Governance performance

Previous literature suggests that corporate governance mechanisms mediate the effect of ESG disclosure regulation. For example, Hummel and Rötzel (2019, p. 209) find that high reporting incentives, as measured by (among other things) the level of corporate governance, mitigate the positive effect of the UK Companies Act on disclosure, as these companies tend to have higher levels of voluntary disclosure prior to regulation (see Section 3.1.1). In addition, mandatory ESG disclosure regimes may also have a direct impact on corporate governance mechanisms by encouraging companies to establish internal monitoring mechanisms for their reporting practices (Gulenko, 2023, p. 33). Although limited, the literature provides positive evidence on the governance outcomes of ESG disclosure regulation.

Boamah (2022, pp. 1–2) examines the impact of the UK's GHG disclosure requirement as part of the Companies Act 2006 Regulations 2013 on the number of directors on board committees dealing with sustainability issues. The results show an absolute (relative) increase of 7.8% (2.6%), which suggests an improvement in ESG-related monitoring and compliance.

Aureli, Del Baldo, et al. (2020, pp. 2392–2394) conduct a paradigmatic case study to examine the effects of the NFRD on the reporting strategy and governance practices of the *Biese Group*, an Italian listed company in the wood and glass processing industry. In order to comply with the new reporting requirements, the company has established an internal audit committee, which is found to play an essential role in guiding the company towards more sustainable behavior. The study also finds that normative and mimetic pressures³⁷ have led to improvements in ESG behavior beyond what is required by regulation. However, the study represents the response of a single company to an ESG disclosure mandate, which limits the generalizability of the results (Aureli, Del Baldo, et al., 2020, p. 2401).

3.3.5. Discussion of findings on real effects

In general, previous studies suggest that ESG disclosure requirements have positive real effects, including improvements in overall ESG performance (e.g., Cuomo et al., 2022; Fiechter et al., 2022), reductions in GHG emissions (e.g., Downar et al., 2021; Jouvenot and Krueger, 2019) and other environmental emissions (e.g., Y.-C. Chen et al., 2018; Gramlich and Huang, 2017), increases in workplace safety (e.g., Y.-C. Chen et al., 2018; H. B. Christensen et al., 2017), and improvements in corporate governance practices (e.g., Aureli, Del Baldo, et al., 2020; Boamah, 2022), among others.

Several studies suggest that the effects are driven by benchmarking mechanisms, such as firms' learning from their peers, or firms wanting to keep up with the performance of their competitors (Fiechter et al., 2022, p. 1515;

³⁷ Normative pressure refers to the pressure to conform with certain social values and norms, and mimetic pressures refers to the firms demand to imitate other companies' behavior. (Aureli, Del Baldo, et al., 2020, p. 2395)

Jouvenot and Krueger, 2019, p. 32; Tomar, 2023, p. 453). In addition, stakeholder pressure is likely to drive potential improvements in ESG activities (Downar et al., 2021, p. 1170; Fiechter et al., 2022, p. 1515; Jouvenot and Krueger, 2019, p. 32; She, 2022, p. 401).

The effectiveness of disclosure regulation in driving real changes in corporate activities appears to depend on certain firm characteristics, such as the voluntary level of ESG disclosure or performance prior to regulation (e.g., Cuomo et al., 2022, p. 7; Fiechter et al., 2022, pp. 1517–1518; Jackson et al., 2020, p. 334). In addition, Cuomo et al. (2022, p. 6) and Fiechter et al. (2022, pp. 1515–1516) argue that companies already begin to adopt more sustainable behaviors after legislation is passed, but before the disclosure requirements come into effect, which is reflected in the early measurement of ESG performance improvements. According to H. B. Christensen et al. (2017, p. 3), better accessibility of previously published information already leads to positive effects in ESG performance. This is measured by a positive incremental effect from the inclusion of ESG information in financial reports, which was already disclosed through other venues beforehand.

While Downar et al. (2021, pp. 1139–1141) report positive real effects with no negative impact on financial firm performance, Y.-C. Chen et al. (2018, p. 170) and H. B. Christensen et al. (2017, p. 5) find that improved ESG performance comes at the expense of profitability.

Much of the existing real effects literature focuses on targeted disclosure regimes, such as the impact of GHG reporting requirements on GHG emissions (e.g., Bauckloh et al., 2023; Downar et al., 2021; Jouvenot and Krueger, 2019; Tomar, 2023). Other studies are limited to single countries (e.g., Y.-C. Chen et al., 2018; Dharmapala and Khanna, 2018; Gramlich and Huang, 2017). As discussed above, the observed effects are therefore likely to be specific to the particular regulation and country setting, and most studies do not specifically account for the impact of these differences have on regulatory outcomes (Ernstberger et al., 2021, p. 3). It is therefore questionable to what extent the findings on these targeted disclosure requirements are applicable to the potential impact of broader ESG disclosure requirements, such as those implemented by the EU (H. B. Christensen et al., 2021, p. 1216; Haji et al., 2023, p. 181). My empirical analysis in Chapter 4 contributes to this ongoing debate.

4. Data analysis: Impact of the NFRD on ESG performance

In this empirical part, I examine the long-term effects of mandatory ESG disclosure regimes on firms' ESG performance. While the topic is increasingly studied, comprehensive empirical evidence remains scarce (H. B. Christensen et al., 2021, p. 1213).

4.1. Background and hypothesis development

Previous research has primarily focused on specific disclosure regulations, such as GHG reporting (e.g., Bauckloh

et al., 2023; Matisoff, 2013; Tomar, 2023), or has been limited to specific countries, such as China (e.g., Y.-C. Chen et al., 2018). Studies that focus on comprehensive disclosure regulations, such as the NFRD, have generally examined their short-term or anticipatory effects (e.g., Cuomo et al., 2022; Fiechter et al., 2022). I aim to fill this research gap by assessing the impact of the NFRD on the ESG performance of affected EU companies. The NFRD was transposed into the national laws of EU member states by December 2016, and applied from 2017 for reports published in 2018 (see Section 2.2.1 for more details).³⁸ In addition to increasing transparency, the NFRD also aims to positively influence companies' ESG behavior (European Parliament and Council, 2014, p. 1, Recital 3).

Previous studies indicate positive anticipatory or short-term effects of the NFRD on ESG performance. For example, Fiechter et al. (2022, pp. 1513–1514) provide evidence of a higher increase in ESG scores both after the announcement of the regulation and in the first year after the NFRD comes into effect. Cuomo et al. (2022, p. 6) also show that firms improve their environmental and social scores in the four years following the enactment of the NFRD (from 2015 to 2018). Similarly, Cicchiello et al. (2023, p. 1125) show that EU companies improve their overall ESG scores in the period from 2018 to 2020. Based on these findings, I derive the following null hypothesis, which I expect to reject through my empirical analysis:

H₀: EU companies subject to the NFRD do not show greater improvements in ESG scores compared to unregulated companies.

4.2. Data and research design

This section begins with an overview of the data and sample used in the analysis (Section 4.2.1), continues with descriptive and exploratory statistics (Section 4.2.2), and concludes with a detailed description of the research design used (Section 4.2.3).

4.2.1. Data and sample

Data collection: In line with related literature (e.g., Cuomo et al., 2022; Fiechter et al., 2022; Jackson et al., 2020; Ottenstein et al., 2022), I obtain ESG data from the LSEG ESG database (formerly Refinitiv ASSET4³⁹). In addition to its widespread use in academia, LSEG ESG provides both detailed and broad measures of companies' ESG activities. Company data is collected on an ongoing basis from sources including annual and ESG reports, company websites, and news sources. It covers over 630 ESG metrics, updated weekly, with ESG scores available for over

³⁸ Greece deviates from this requirement in that the country has mandated compliance with the disclosure requirements one year earlier than other EU member states (CSR Europe and GRI, 2017, p. 33, end note i).

³⁹ Refinitiv has been acquired by the London Stock Exchange Group (LSEG) in 2021. (LSEG, 2023a)

15,000 companies worldwide. (LSEG, 2022, pp. 3–6) Financial company data is sourced from the comprehensive LSEG Worldscope Fundamentals database (LSEG, 2023b).

Sample: My sample of treatment and control firms is based on the sample of Fiechter et al. (2022), who study the yearly treatment effect of the NFRD over the period 2011 to 2018.⁴⁰ Based on the ISINs⁴¹ of their final matched sample, which are published as part of the online supplements to the paper (Chicago Booth, 2023), I was able to extract relevant and current company information from the LSEG databases. The treatment group consists of large public EU companies subject to the NFRD.⁴² The control group is a propensity score matched sample of US companies.⁴³ Propensity score matching (PSM) is a statistical matching technique used to create a control group that is similar in characteristics (such as industry classification, prior disclosure levels, etc.) to the treatment group (Bell, 2021, p. 12; Rosenbaum and Rubin, 1983, p. 48). PSM is often used by researchers seeking to isolate the effects of a policy change from other confounding variables (e.g., Fiechter et al., 2022, p. 1506; Ioannou and Serafeim, 2017, p. 4).

US companies are an appropriate control group for two main reasons: first, no comprehensive ESG disclosure regulations were introduced in the US during the period under study (Cuomo et al., 2022, pp. 14–15; Fiechter et al., 2022, p. 1506) and second, LSEG ESG has a broad coverage of ESG data for US companies (LSEG, 2022, p. 7). Since some ISIN numbers of the original Fiechter et al. (2022) sample are no longer available (e.g. due to delisting), I eliminate these observations. I also eliminate duplicate entries.⁴⁴ This results in a final sample of 685 firms (528 treatment and 157 control firms) over an observation period of 8 to 12 years, leading to 7947 firm-year observations. I use the statistical software R (version 4.3.2) to conduct the empirical analysis.

Sample period: The analysis includes data from 2011 to 2022, divided into two phases: pre-directive (2011 to 2016) and post-directive (2017 to 2022). The analysis focuses specifically on the impact of the NFRD after 2017, which is consistent with the start of the new reporting requirements. This differs from other studies that assess the anticipatory

effects after 2014, i.e. after the adoption but before the entry into force of the NFRD (e.g., Cuomo et al., 2022, p. 6).

Sample distribution: Table 1 presents the sample distribution. The country distribution in Panel A shows that the majority of firms in the treatment group are incorporated in the UK (24.7%), France (10.1%), and Germany (8.6%).⁴⁵ This distribution is consistent with the original Fiechter et al. (2022) sample. All firms in the control group are incorporated in the US. Panel B shows the industry distribution of the treatment and control groups. Firms in the treatment group are predominantly in the industrial (22.3%), consumer cyclical (16.5%), and financial (16.3%) sectors. The distribution of the control group differs slightly but is overall aligned.

4.2.2. Descriptive and exploratory statistics

Next, I analyze the metric variables (i.e., firm characteristics and ESG scores) of the sample firms in more detail.⁴⁶ I visualize the distribution by creating histograms and box plots for each variable, which allows me to identify skewness and outliers (see Panel A and Panel B of Appendix 4, respectively). First, I perform logarithmic transformations to address high skewness of some variables. Second, I winsorize the variables to limit the influence of outliers. This means that extreme outliers are replaced by less extreme values (R Documentation, 2024). In line with related literature (e.g., Cicchiello et al., 2023, p. 1124; Fiechter et al., 2022, p. 1549), I replace the lower and upper 1% of the data (i.e., the values outside the 1st and 99th percentiles) of each variable. Appendix 5 presents histograms and box plots of the variables after data cleaning. The largest outliers have been winsorized, resulting in more symmetric and potentially more normal distributions of the variables.

Table 2 shows the means and standard deviations after data cleaning for the full sample and for the treatment and control groups separately, both for the entire observation period and separately for the pre- and post-directive periods. For the entire period (2011 to 2022), the average ESG score for the full sample is 60%, while the average score for the treatment (control) group is 61% (56%). Table 2 also shows that the average ESG score for the full sample increases from 55% to 65% from the pre-directive period (2011 to 2016) to the post-directive period (2017 to 2022). For the EU (US) companies, the average ESG score increases from 56% (52%) to 67% (61%) from the pre- to the post-NFRD period. This is a first indication that EU firms have a higher increase in ESG scores than US firms, which is later statistically tested.

In the full sample, firms have on average an ROA of 0.05, a firm size of 16.34 and a leverage of 0.65. The scores and firm characteristics are also consistent with related studies examining the impact of the NFRD on ESG performance (Cuomo et al., 2022, p. 7; Fiechter et al., 2022, pp. 1508–1509). Appendix 6 also presents the more detailed parametric measures of location and dispersion before and after data

⁴⁰ For a detailed description of their sample selection process, refer to Fiechter et al. (2022, p. 1507)

⁴¹ The International Securities Identification Number is an international standard to uniquely identify securities.

⁴² Greece implemented NFRD disclosure requirements for companies in 2017, one year ahead of other EU states (CSR Europe and GRI, 2017, p. 33, end note i). Although Cuomo et al. (2022, p. 6) exclude Greek firms for this reason, I include them to align with the matched sample of Fiechter et al. (2022). Moreover, Greek firms account for only 1.2% of my total sample.

⁴³ Fiechter et al. (2022) base the propensity score matching on several variables, including ESG activities and reporting, industry classification, firm size, leverage, and return on assets, among others. For a detailed description, refer to Fiechter et al. (2022, p. 1506).

⁴⁴ As part of the matching procedure, Fiechter et al. (2022, p. 1506, footnote 8) have allowed for replacement in their control group. This means the same firm may be included in the control group multiple times. However, since R cannot deal with double entries, I reduce my control group to include each firm only once.

⁴⁵ Percentages refer to the entire sample.

⁴⁶ All variables are defined in Appendix 3.

Table 1: Sample distribution across countries and industries (source: own representation)

Panel A. Country distribution		
	n obs.	% obs.
Austria	132	1.6
Belgium	192	2.3
Czech Republic	36	0.4
Denmark	216	2.6
Finland	264	3.2
France	828	10.1
Germany	708	8.6
Greece	96	1.2
Hungary	36	0.4
Ireland	192	2.3
Italy	312	3.8
Luxembourg	48	0.6
Netherlands	288	3.5
Poland	192	2.3
Portugal	60	0.7
Spain	372	4.5
Sweden	336	4.1
United Kingdom	2028	24.7
United States	1884	22.9
<i>TOTAL</i>	<i>8220</i>	<i>100.0</i>

This table presents the country distribution of the entire sample. Country of incorporation is derived from the LSEG *COINN* variable. All control group firms are incorporated in the US.

Panel B. Industry distribution				
	EU Firms		US Firms	
	n obs.	% obs.	n obs.	% obs.
Academic & Educational Services	0	0.0	24	1.3
Basic Materials	708	11.2	180	9.6
Consumer Cyclicals	1044	16.5	300	15.9
Consumer Non-Cyclicals	456	7.2	96	5.1
Energy	276	4.4	60	3.2
Financials	1032	16.3	384	20.4
Healthcare	336	5.3	108	5.7
Industrials	1416	22.3	348	18.5
Real Estate	84	1.3	84	4.5
Technology	648	10.2	252	13.4
Utilities	336	5.3	48	2.5
<i>TOTAL</i>	<i>6336</i>	<i>100.0</i>	<i>1884</i>	<i>100.0</i>

This table presents the industry distribution of the treatment and the control group. Industry classifications is derived from the LSEG *TR1N* variable to cluster companies by business sector.

Table 2: Summary statistics clustered by period and group (source: own representation)

	EU Firms		US Firms		Full sample	
	Mean	SD	Mean	SD	Mean	SD
Entire period						
ESG Score	61.36	17.86	56.17	18.65	60.19	18.17
Env. Pillar Score	60.93	24.15	50.67	26.99	58.61	25.19
Social Pillar Score	64.07	21.19	58.03	21.47	62.70	21.40
ROA	0.04	0.06	0.05	0.07	0.05	0.06
Firm Size	16.30	2.01	16.47	1.55	16.34	1.91
Leverage	0.65	0.21	0.66	0.21	0.65	0.21
PRE period						
ESG Score	56.41	18.64	52.16	19.18	55.43	18.85
Env. Pillar Score	57.69	25.37	47.81	27.44	55.43	26.19
Social Pillar Score	58.02	22.29	54.02	21.95	57.11	22.28
ROA	0.04	0.06	0.05	0.06	0.05	0.06
Firm Size	16.15	2.02	16.29	1.55	16.19	1.92
Leverage	0.64	0.21	0.65	0.22	0.65	0.21
POST period						
ESG Score	66.63	15.33	60.60	16.99	65.29	15.91
Env. Pillar Score	64.37	22.28	53.82	26.14	62.02	23.60
Social Pillar Score	70.49	17.85	62.45	20.03	68.70	18.65
ROA	0.04	0.06	0.05	0.07	0.04	0.06
Firm Size	16.44	1.98	16.65	1.53	16.49	1.89
Leverage	0.65	0.21	0.67	0.20	0.66	0.21

This table reports means and standard deviations for EU treatment firms, US control firms, and the full sample. Information is provided for the entire sample period (2011 to 2022), the pre-directive period (2011 to 2016), and the post-directive period (2017 to 2022). All variables are defined in Appendix 3.

cleaning, and Appendix 7 shows the corresponding tables for the non-parametric measures. Again, the tables indicate that the data cleaning process has effectively corrected for extreme outliers and skewness, making the dataset more robust for further analysis.

In addition to examining each variable individually as described above, I also analyze them in relation to each other as part of exploratory statistics. The correlation heatmaps provide a first indication of the relationships between the metric variables (see Appendix 8, Panel A). First, the visualization of the correlations between the variables shows that there is multicollinearity between the accounting variables, which may limit the reliability of the results. However, the pairwise correlations between ROA, firm size, and leverage range from -0.35 to 0.37, indicating rather weak relationships (Schober et al., 2018, p. 1765). This level of correlation does not raise significant multicollinearity concerns, suggesting that these variables may be included in the estimation models. All pairwise correlations are significant at $p < .05$. I also use bivariate scatterplots to visualize the relationship between firm

characteristics and ESG scores (see Appendix 8, Panel B). The plots suggest that firm size and leverage are positively correlated with ESG scores, while ROA has a weak negative relationship. However, the data points are widely scattered around the line, indicating rather weak relationships.

4.2.3. Research design

Before running the regression model, I further test whether the assumption of normal distribution is violated. The Jarque Bera test provides evidence of non-normal distribution for all variables. Visual inspection using quantile-quantile (QQ) plots shows that while the data mostly aligns with the expected normal line in the central quantiles, there are notable deviations, especially in the tails (see Appendix 9). This deviation from the normal distribution could affect the robustness of the regression results, which is discussed in the limitations in Section 4.6.

Model description: I use a difference-in-difference (DID) analysis to test my hypothesis. The DID is a quasi-experimental approach commonly used to evaluate the impact of an ex-

ogenous shock, such as the introduction of a new regulation (e.g., Cuomo et al., 2022; Fiechter et al., 2022; Ioannou and Serafeim, 2017; Ottenstein et al., 2022; Tomar, 2023). The methodology compares changes in the dependent variable before and after the implementation of a regulation between the treatment group and the control group of unaffected companies. The difference in the change in the outcome variable between the two groups is the DID estimate of the average effect of the regulation. (Greene, 2012, pp. 195–197) I control for industry membership, country of incorporation, and other observable firm characteristics.

To test my hypothesis, I first use a pooled Ordinary Least Squares (OLS) regression model as a baseline model (Greene, 2012, pp. 389–390) to estimate the following DID equation:

$$\begin{aligned} ESG\ Score_{it} = & \beta_0 + \beta_1 (Post * Treatment)_{it} \\ & + \beta_2 Post_t + \beta_3 Treatment_i + \beta_4 ROA_{it} \\ & + \beta_5 Leverage_{it} + \beta_6 Firm\ Size_{it} \\ & + \beta_7 Industry_i + \beta_8 Country_i + \varepsilon_{it} \end{aligned}$$

Variable description: The subscript i denotes the individual company and the subscript t denotes the year, meaning that $ESG\ Score_{it}$ is the outcome variable for company i in year t . I choose LSEG's ESG score, which is a comprehensive measure of a company's ESG performance, commitment, and effectiveness (LSEG, 2022, p. 8).⁴⁷ I choose this broad measure as my dependent variable for two reasons: first, a comprehensive measure is more appropriate given my diverse set of companies, which represents the wide range of companies affected by the NFRD (Fiechter et al., 2022, pp. 1512–1513). Second, LSEG's ESG score has been widely used in relevant related literature (e.g., Cicchiello et al., 2023, p. 1125; Cuomo et al., 2022, p. 6).

The binary variable *Post* indicates the observation period and equals 1 for all firm-year observation after the directive's coming into force, i.e. from 2017 to 2022. The binary variable *Treatment* is 1 if the company belongs to the EU treatment group, and 0 if it belongs to the US control group. The interaction term *Post * Treatment* is equal to 1 if entity i is affected by the NFRD in year t (i.e. for EU companies in the post-directive period), and 0 otherwise. The coefficient estimator of interest is β_1 , which captures the incremental effect of the disclosure regulation on treated firms' ESG performance, relative to the control group. In line with related literature (Cicchiello et al., 2023; Cuomo et al., 2022; Fiechter et al., 2022; Jackson et al., 2020; Krueger et al., 2023), I control for the following firm-level characteristics that are likely

to influence ESG Scores: firm profitability (*ROA*) as the net income scaled by total assets, *Leverage* as the total liabilities to total assets ratio, and *Firm Size* as the natural logarithm of total assets.⁴⁸ High profitability and large company size are likely to have a positive impact on ESG performance due to the availability of resources to invest in ESG activities, while high leverage implies greater financial constraints, which could have a negative impact on ESG scores (e.g., Cicchiello et al., 2023, p. 1125; Jackson et al., 2020, p. 330). I also include *Industry* and *Country* dummies to control for industry- and country-specific differences that potentially influence ESG performance. ε is the unobserved error term. All variables are defined in Appendix 3.

4.3. Results: Positive impact on ESG performance

Appendix 10 presents the results of the pooled OLS regression analysis. The interaction term *Post * Treatment*, which captures the differential effect of the NFRD for the treatment group, has a significantly positive coefficient estimate of 2.36 (p -value < .01). This implies that affected EU companies improve their ESG scores on average by an additional 2.36 percentage points relative to the US control group, which rejects the null hypothesis H_0 . The estimate translates into a 4.2% increase in ESG scores relative to the mean pre-directive ESG score of treatment firms.⁴⁹ These results are aligned with Fiechter et al. (2022, p. 1514), who report a 5.5% increase in ESG scores in 2018 relative to their baseline year of 2013. *Post* has a significantly positive coefficient of 6.08 (p -value < .001), indicating that ESG scores increase on average after the regulation for both the treatment and the control group. *Treatment* also has a significantly positive coefficient of 7.33 (p -value < .001), indicating that the treatment group has higher ESG scores on average than the control group. As for the control variables, *ROA* and *Firm Size* have significantly positive coefficients of 11.77 (p -value < .001) and 6.25 (p -value < .001), respectively. This suggests that more profitable companies and larger companies tend to have higher ESG scores. While *Leverage* has a positive coefficient of 1.66, its statistical significance is only at the 10% level. This provides only weak evidence for the relationship between leverage and ESG scores.

The coefficients for each country represent the average difference in ESG scores for companies in that country compared to the reference country, which is the US in my case. For example, a negative coefficient of 5.42 (p -value < .001) for Austria indicates a lower average ESG score for companies in Austria compared to the US. The coefficients for each industry represent the average difference in ESG scores for

⁴⁷ LSEG's ESG score is constructed from over 630 firm-level data points, of which the 186 most relevant are selected for each industry. Based on these measures, ten category scores are calculated, which contribute to the assessment of the environmental, social and governance pillar scores and ultimately to the company's overall ESG score. The ESG score ranges from 0 to 100, with 100 being the maximum score. A detailed description of the scoring process is available at LSEG (2022).

⁴⁸ I limit the number of control variables to the most commonly used ones, in contrast to the more comprehensive analyses in the aforementioned studies.

⁴⁹ $4.2\% = 2.36/56.41$, where 2.36 is the regression coefficient β_1 for the interaction term *Post * Treatment* in column (1) of Appendix 10, and 56.41 is the mean ESG Score for treatment firms in the pre-directive period in Table 2.

companies in that industry compared to the reference industry, Academic & Educational Services. For example, companies in the Real Estate sector have on average higher ESG scores, with a positive coefficient of 9.98 (p -value < .01).

In terms of model fit, the R-squared value of 0.426 indicates that the model explains 42.6% of the variance in the ESG scores. The F-statistic indicates that the overall model is highly significant with $F(33, 7913) = 177.83$, $p < .001$.

I perform several diagnostic tests to assess the reliability of the results by analyzing the residual errors for normal distribution, heteroskedasticity, and serial correlation (see Appendix 11). To test whether the residuals are normally distributed, I use the Anderson-Darling normality test. The small p -value (< .001) indicates that the residuals are not normally distributed. I also use a QQ plot as a graphical tool to inspect the distribution of the residuals in my dataset: while most of the points lie along the line, there are some deviations in the tails (see Appendix 11, Panel A). Next, the Breusch-Godfrey test suggests the presence of serial correlation (p -value < .001) (see Appendix 11, Panel B). Finally, the studentized Breusch-Pagan test suggests the presence of heteroskedasticity (p -value < .001). The residuals plot supports this finding graphically (see Appendix 11, Panel C).

To address the issues of heteroskedasticity and serial correlation in the pooled OLS regression model, I compute robust standard errors to enhance the reliability of the estimates (Greene, 2012, pp. 390–392). The results from the Sandwich Estimator are presented in column (3) of Appendix 10. Importantly, the positive coefficient of the interaction term ($Post * Treatment$) remains statistically significant (p -value < .01). Overall, the application of the *SandwichEstimator* provides additional evidence for the main finding of my pooled OLS model, namely that the NFRD has a positive impact on ESG performance of affected EU companies, relative to the US control group.

4.4. Robustness checks and alternative analyses

Robustness checks: To test the robustness of the results, I rerun my analysis using an individual fixed effect model⁵⁰, which controls for individual entity-specific fixed effects:

$$ESG\ Score_{it} = \alpha_i + \beta_1 (Post * Treatment)_{it} + \beta_2 Post_{it} + \beta_4 ROA_{it} + \beta_5 Leverage_{it} + \beta_6 Firm\ Size_{it} + \varepsilon_{it}$$

Where α_i captures all the individual fixed effects, thereby controlling for time-invariant unobservable differences in company characteristics. Therefore, the time-invariant variables *Industry*, *Country*, and *Treatment* from the baseline pooling model are not specifically included in the equation. Although the fixed effect model reduces omitted variable

bias, it cannot estimate the coefficients of the time-invariant variables. (Greene, 2012, pp. 399–400)

I run the same diagnostic test as for the pooled OLS model above. To address the issue of heteroskedasticity and serial correlation, I again use the Sandwich Estimator to compute robust standard errors. As tabulated in Appendix 12, the coefficient on the interaction term $Post * Treatment$ remains positive and statistically significant, with a robust estimate of 2.26 (p -value < .01). This supports the previous finding that the NFRD has a positive impact on the ESG performance of affected companies in the post-regulation period. $Post$ is also significantly positive with an estimate of 6.38 (p -value < .001), indicating that all sample companies, on average, increase their ESG performance after the directive comes into effect, relative to the pre-directive period. Among the control variables, only *Firm Size* is significantly positive, again suggesting that larger companies have higher ESG scores on average. *ROA* and *Leverage* have positive but insignificant coefficients.

Alternative analyses: In line with related literature (Cuomo et al., 2022, p. 6; Fiechter et al., 2022, p. 1513), I conduct additional analyses using alternative dependent variables, namely LSEG's Environmental Pillar Score and the Social Pillar Score. As presented in Appendix 13, the results of the pooled OLS model indicate that the coefficient of the interaction term for the Environmental Pillar Score is positive, but insignificant. Appendix 14 shows that the coefficient of the Social Pillar Score is significantly positive at 4.65 (p -value < .001).

4.5. Discussion of empirical results

My analysis provides evidence against the null hypothesis H_0 by showing that the NFRD encourages companies to adopt more sustainable behavior. Affected EU companies have a higher increase in ESG scores following the NFRD relative to unregulated firms. This finding is also consistent with previous research (Cicchello et al., 2023; Cuomo et al., 2022; Fiechter et al., 2022). The improvement seems to be mainly driven by better social performance, which is consistent with Fiechter et al. (2022, p. 1513), but differs from Cuomo et al. (2022, p. 14), who observed significant improvements in both social and environmental aspects. Among the control variables, only *Firm Size* has a significantly positive estimate in both models, suggesting that larger companies have higher ESG scores on average. As discussed above, this is reasonable because these companies are likely to have more resources to invest in ESG-related activities (e.g., Cicchiello et al., 2023, p. 1125; Jackson et al., 2020, p. 330).

Two primary mechanisms, stakeholder pressure and peer benchmarking, are likely to drive the observed ESG improvements following NFRD implementation. First, stakeholders may respond to the newly disclosed information by pressuring reporting firms to improve their ESG performance, which may lead to the adoption of more sustainable corporate behavior (e.g., Y.-C. Chen et al., 2018, p. 171; Downar et al., 2021, p. 1170; Jouvenot and Krueger, 2019, p. 32). Second,

⁵⁰ In line with Ernstberger et al. (2021, p. 17), I use the Hausman test to decide whether the use of a fixed effect or a random effect model is more appropriate for my data (Greene, 2012, pp. 419-420). The untabulated results support the use of a fixed effects model (p -value < .05).

the increased transparency and availability of ESG information allows companies to benchmark their own ESG performance against their peers. The potential to learn from them and the desire to outperform competitors can lead to better ESG performance (e.g., Jouvenot and Krueger, 2019, p. 32; Tomar, 2023, p. 454). In addition, the relative ease of implementing social initiatives compared to more complex environmental initiatives could be a possible reason for the more significant increase in social activities observed (Fiechter et al., 2022, p. 1521).

My study contributes to the understanding of the real effects of ESG disclosure regulation. Previous studies have largely focused on targeted ESG regulations, such as the US GHG Reporting Program (Bauckloh et al., 2023; Tomar, 2023). These regulations are often specifically designed to encourage a particular corporate behavior, such as reducing GHG emissions. In contrast, there is limited research on broader ESG regulations, such as the NFRD, that apply across industries and countries. Previous research on the NFRD tends to focus on anticipatory or short-term effects (Cuomo et al., 2022; Fiechter et al., 2022). My findings contribute to the debate on the effectiveness of such comprehensive regulations by demonstrating positive long-term effects on ESG performance over six years post-regulation.

My findings also have practical implications by informing policymakers and companies (potentially) subject to reporting requirements about the effectiveness of comprehensive ESG disclosure mandates. The success of the NFRD in encouraging ESG activity under a comply-or-explain approach is relevant beyond the EU, particularly for countries considering similar ESG disclosure requirements. In light of the forthcoming CSRD, which affects approximately 50,000 companies across Europe, my research also supports the EU's move towards stricter regulation and its commitment to an enhanced ESG regulatory framework.

4.6. Limitations of empirical analysis and suggestions for future research

My empirical analysis is subject to several limitations.

Differences in national legislation: While the NFRD is a supranational law, member states have some leeway in their national transposition. This results, for example, in different assurance requirements, penalties, disclosure formats, and scope of companies affected across countries (CSR Europe and GRI, 2017, p. 10). To fully understand the impact of the NFRD, future studies could investigate how these regulatory differences affect ESG performance.

Influence of the institutional environment and cultural aspects: I do not specifically explore the implications of differences in countries' institutional environments on ESG performance. However, the strength of the legal environment and enforcement mechanisms, as well as cross-cultural differences across EU member states, are likely to influence the effectiveness of the regulation (Krueger et al., 2023, pp. 44–45; Mittelbach-Hörmanseder et al., 2021, p. 335). Future research could further explore this topic.

Statistical limitations: The non-normal distribution of the variables and the residuals may limit the reliability of the results. Although the results remain significant when running different models and using robust standard errors, further robustness checks could include alternative control groups (Ioannou & Serafeim, 2017, pp. 25–26), alternative standard error clustering techniques (Fiechter et al., 2022, p. 1509) or alternative fixed effects (Ernstberger et al., 2021, p. 23).

Limitations of LSEG's ESG Score: While useful to capture the impact of the NFRD on the broad set of companies affected, LSEG's ESG score has its own limitations. I cannot rule out whether the methodological approach for calculating the score has changed during my sample period, which could affect ESG scores and hence the results. In addition, the score is based on public reporting, so improvements in ESG scores may reflect changes in reporting rather than actual changes in ESG activity. For example, companies may have adopted ESG-related activities prior to regulation, but only start disclosing the information after being subject to mandatory reporting requirements (Jackson et al., 2020, p. 330). Future research could explore more objective measures, such as GHG emissions. However, these measures may lack comparability across the different types of companies covered by the NFRD.

Isolating the effects of regulation: It is difficult to disentangle the effects of the NFRD from broader ESG trends or other regulatory changes. For example, ESG-related incidents, other market-wide shocks, or the introduction of related regulation affecting EU firms differently than US firms may explain the relative improvement in ESG performance, rather than the shift from voluntary to mandatory disclosure. (H. B. Christensen et al., 2021, p. 1213; Fiechter et al., 2022, p. 1542; Krueger et al., 2023, p. 32) Although these problems with causal inference are inherent in studies of regulatory change (Leuz & Wysocki, 2016, pp. 535–538), future research could focus on isolating firm responses to ESG disclosure regulation from other confounding effects.

Underlying mechanisms and drivers: While I argue that stakeholder pressure and peer benchmarking likely influence the positive impact of the NFRD on ESG performance, I do not specifically examine how these mechanisms influence the observed effect. Further studies could examine their role in detail. In addition, it may be interesting to further investigate why the improvements in ESG scores seem to be driven by better social performance, as opposed to better environmental performance.

5. Conclusion

This thesis advances the understanding of the impact of mandatory ESG disclosure regulations on companies by employing a threefold approach: a global analysis of ESG disclosure regulations, a literature review of their impact on reporting companies, and an empirical analysis of the effectiveness of the EU NFRD in promoting sustainable corporate behavior.

First, I find that disclosure requirements differ substantially across countries with respect to several dimensions (see

Chapter 2). These include the type of issuer (government versus stock exchange), the scope of content requirements and firms covered, and the basis of reporting. However, regulators around the world have recognized the need for standardization. While this may lead to greater convergence of reporting requirements, institutional and cultural differences between countries are likely to influence the effectiveness of regulations.

Second, the literature review in Chapter 3 suggests that ESG disclosure regulation has a positive impact on firms' disclosure practices (as measured by increased disclosure quantity and quality), although other studies provide no supporting evidence. Results on the financial effects of ESG disclosure regulation, including capital market and firm profitability effects, are also mixed. ESG mandates seem to lead to better ESG performance of companies, although many of these studies focus on targeted regulation. The heterogeneity of results may be due to differences in reporting requirements and institutional contexts, as identified in Chapter 2 of the thesis. Future research needs to account for these variations in order to draw precise implications from empirical results on the consequences of ESG disclosure mandates.

Third, I provide empirical evidence that the EU's NFRD, a comprehensive and widespread ESG disclosure mandate, has a positive impact on sustainable corporate behavior (see Chapter 4). In particular, the results from the DID analysis suggest that affected companies improve their ESG performance following the coming into force of the directive, relative to a matched sample of unregulated US companies. This effect seems to be driven by improvements in the social dimension. My analysis differs from most research that focuses on early effects, as I observe changes that materialize in the six years after companies have to comply with the reporting requirements. The results are of particular interest to regulators seeking to understand the effectiveness of disclosure mandates, as they suggest that the shift from voluntary to mandatory ESG disclosure is having the desired positive real effects.

Future studies could build on these findings by examining the impact of the recently adopted CSRD, particularly its extension to listed SMEs. Given that these firms often face knowledge and resource constraints (European Commission, 2022a, p. 81), it is important to understand how SMEs navigate the extensive ESG disclosure requirements. Although the CSRD affects a wider range of companies, its overall impact is still relatively limited. As of 2022, there are approximately 24 million active SMEs in the EU, accounting for more than 99% of businesses (European Commission, 2023b, p. 6, 2024) and responsible for over 60% of industrial CO₂ and GHG emissions in the EU (European Commission, 2022a, p. 16). With a total of approximately 50,000 EU firms subject to the CSRD, mandatory ESG disclosure requirements still affect only a fraction of companies. While most existing research focuses on the firm-level effects of these requirements, the broader, economy-wide implications of ESG disclosure (if any) remain an important but largely unexplored area for future research.

References

- Accountancy Europe. (2020). Towards reliable non-financial information across Europe. Retrieved January 4, 2024, from https://www.accountancyeurope.eu/wp-content/uploads/2022/12/Accountancy-Europe-NFI-assurance-practice_factsheet.pdf
- Agostini, M., Costa, E., & Korca, B. (2022). Non-Financial Disclosure and Corporate Financial Performance Under Directive 2014/95/EU: Evidence from Italian Listed Companies. *Accounting in Europe*, 19(1), 78–109. <https://doi.org/10.1080/17449480.2021.1979610>
- Amel-Zadeh, A., & Serafeim, G. (2018). Why and How Investors Use ESG Information: Evidence from a Global Survey. *Financial Analysts Journal*, 74(3), 87–103. <https://doi.org/10.2139/ssrn.2925310>
- Antal, A. B., & Sobczak, A. (2007). Corporate Social Responsibility in France. *Business & Society*, 46(1), 9–32. <https://doi.org/10.1177/0007650306293391>
- Arvidsson, S., & Dumay, J. (2022). Corporate ESG reporting quantity, quality and performance: Where to now for environmental policy and practice? *Business Strategy and the Environment*, 31(3), 1091–1110. <https://doi.org/10.1002/bse.2937>
- Aureli, S., Del Baldo, M., Lombardi, R., & Nappo, F. (2020). Nonfinancial reporting regulation and challenges in sustainability disclosure and corporate governance practices. *Business Strategy and the Environment*, 29(6), 2392–2403. <https://doi.org/10.1002/bse.2509>
- Aureli, S., Salvatori, F., & Magnaghi, E. (2020). A Country-Comparative Analysis of the Transposition of the EU Non-Financial Directive: An Institutional Approach. *Accounting, Economics, and Law*, 10(2), 1–30. <https://doi.org/10.1515/ael-2018-0047>
- Australian Government. (2007). National Greenhouse and Energy Reporting Act 2007 (NGER). Retrieved December 19, 2023, from <https://www.legislation.gov.au/C2007A00175/latest/text>
- Australian Government. (2018). Modern Slavery Act 2018. Retrieved November 17, 2023, from <https://www.legislation.gov.au/Details/C2018A00153>
- Australian Government. (2023). Climate-related financial disclosure: Consultation paper. Retrieved November 17, 2023, from <https://treasury.gov.au/sites/default/files/2023-06/c2023-402245.pdf>
- Barth, M. E., Cahan, S. F., Chen, L., & Venter, E. R. (2017). The economic consequences associated with integrated report quality: Capital market and real effects. *Accounting, Organizations and Society*, 62, 43–64. <https://doi.org/10.1016/j.aos.2017.08.005>
- Bauckloh, T., Klein, C., Pioch, T., & Schiemann, F. (2023). Under Pressure? The Link Between Mandatory Climate Reporting and Firms' Carbon Performance. *Organization & Environment*, 36(1), 126–149. <https://doi.org/10.1177/10860266221083340>
- Bell, N. (2021). *Determinants and consequences of CSR disclosure: normative and empirical evidence* [Research paper]. Technical University of Munich. Retrieved July 5, 2023, from <https://mediatum.ub.tum.de/doc/1586315/1586315.pdf>
- Birkey, R. N., Guidry, R. P., Islam, M. A., & Patten, D. M. (2018). Mandated Social Disclosure: An Analysis of the Response to the California Transparency in Supply Chains Act of 2010. *152(3)*, 827–841. Retrieved October 22, 2023, from https://www.jstor-org.eaccess.tum.edu/stable/pdf/45022770.pdf?refreqid=fastly-default%3A7b729e088b13a3892ffe9b729ca4fef0&ab_segments=&origin=&initiator=&acceptTC=1
- Blaschke, M., Frank, M., & Friedl, G. (2023). *Nachhaltigkeitskriterien im Mittelstand: Werden aktuelle Regulatorien im Bereich Nachhaltigkeit zu Finanzierungsschwierigkeiten im deutschen Mittelstand führen?* [Research study]. Hanns-Seidel-Stiftung e.V.
- Boamah, E. O. (2022). Mandatory carbon disclosure and green committees. *Economics Letters*, 219, 1–4. <https://doi.org/10.1016/j.econlet.2022.110767>
- Bundestag. (2017). Gesetz zur Stärkung der nichtfinanziellen Berichterstattung der Unternehmen in ihren Lage- und Konzernlageberichten (CSR-Richtlinie-Umsetzungsgesetz). Retrieved November 20, 2023, from https://www.bmj.de/SharedDocs/Downloads/DE/Gesetzgebung/BGBl/BGBl_CSR-RiLi_UmsetzungsgesetzG.pdf?__blob=publicationFile&v=3

- Bundestag. (2021). Gesetz über die unternehmerischen Sorgfaltspflichten in Lieferketten (Lieferkettensorgfaltspflichtengesetz). Retrieved August 30, 2023, from https://www.bgbl.de/xaver/bgbl/text.xav?SID=&tf=xaver.component.Text_0&toctf=&qmf=&hlf=xaver.component.Hitlist_0&bk=bgbl&start=%2F%2F%5B%40node_id%3D%27940343%27%5D&skin=pdf&tlevel=-2&nohist=1&siinst=35BF2F71
- Bursa Malaysia. (2026). Amendments on various enhancements to the listing requirements of Bursa Malaysia Securities Berhad. Retrieved November 18, 2023, from https://www.bursamalaysia.com/site/s/5bb54be15f36ca0af339077a/content_entry5ce3b50239fba2627b2864be/5ce3c2325b711a1437a700ce/files/regulation_rules_main_market_archives_LRMBSE_Listing_fees_141206.pdf?1570701443
- Bursa Malaysia. (2015a). ACE LR: Appendix 9C - Contents of annual report. Retrieved November 19, 2023, from <https://asklisting.bursamalaysia.com:8443/ACE/Pages/AceAppendix9C.aspx>
- Bursa Malaysia. (2015b). Main LR: Appendix 9C - Contents of annual report. Retrieved November 19, 2023, from <https://asklisting.bursamalaysia.com:8443/MainLR/Pages/MainAppendix9C.aspx>
- Bursa Malaysia. (2015c). Main Market Listing Requirements: Amendments relating to Sustainability Statement in annual reports. Retrieved November 19, 2023, from https://www.bursamalaysia.com/site/s/5bb54be15f36ca0af339077a/content_entry5ce3b5005b711a1764454c1a/5ce3c81c39fba2619023d12d/files/Appendix_1_MainLR_Sustainability_final_7Oct2015_.pdf?1570701455
- Bursa Malaysia. (2022). Sustainability Reporting Guide 2022 (3rd Edition). Retrieved November 18, 2023, from https://bursaacademy.s3.ap-southeast-1.amazonaws.com/Print-PDF/Sustainability%20Reporting%20Guide%202022_FINAL.pdf
- Business & Human Rights Resource Centre. (2017). France's Duty of Vigilance Law. Retrieved August 22, 2023, from <https://www.business-humanrights.org/en/big-issues/corporate-legal-accountability/frances-duty-of-vigilance-law/>
- California Government. (2010). The California Transparency in Supply Chains Act of 2010 (CTSCA). Retrieved November 14, 2023, from https://oag.ca.gov/sites/all/files/agweb/pdfs/cybersafety/sb_657_bill_ch556.pdf
- California Government. (2023a). SB-253 Climate Corporate Data Accountability Act. Retrieved August 16, 2023, from https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB253
- California Government. (2023b). SB-261 Greenhouse gases: climate-related financial risk. Retrieved August 16, 2023, from https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB261
- Chauvey, J.-N., Giordano-Spring, S., Cho, C. H., & Patten, D. M. (2015). The Normativity and Legitimacy of CSR Disclosure: Evidence from France. *Journal of Business Ethics*, 130(4), 789–803. <https://doi.org/10.1007/s10551-014-2114-y>
- Chen, S., Song, Y., & Gao, P. (2023). Environmental, social, and governance (ESG) performance and financial outcomes: Analyzing the impact of ESG on financial performance. *Journal of Environmental Management*, 345, 118829. <https://doi.org/10.1016/j.jenvman.2023.118829>
- Chen, Y.-C., Hung, M., & Wang, Y. (2018). The effect of mandatory CSR disclosure on firm profitability and social externalities: Evidence from China. *Journal of Accounting and Economics*, 65(1), 169–190. <https://doi.org/10.1016/j.jaccoco.2017.11.009>
- Chicago Booth. (2023). Online Supplements and Datasheets - Volume 60. Retrieved December 17, 2023, from <https://www.chicagobooth.edu/research/chookaszian/journal-of-accounting-research/online-supplements-and-datasheets/volume-60>
- Chinese Government. (2014). Environmental Protection Law of the People's Republic of China. Retrieved November 10, 2023, from https://www.gov.cn/zhengce/2014-04/25/content_2666434.htm
- Christensen, D. M., Serafeim, G., & Sikochi, A. (2022). Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings. *The Accounting Review*, 97(1), 147–175. <https://doi.org/10.2308/TAR-2019-0506>
- Christensen, H. B., Floyd, E., Liu, L. Y., & Maffett, M. (2017). The real effects of mandated information on social responsibility in financial reports: Evidence from mine-safety records. *Journal of Accounting and Economics*, 64(2-3), 284–304. <https://doi.org/10.1016/j.jaccoco.2017.08.001>
- Christensen, H. B., Hail, L., & Leuz, C. (2021). Mandatory CSR and sustainability reporting: economic analysis and literature review. *Review of Accounting Studies*, 26(3), 1176–1248. <https://doi.org/10.1007/s11142-021-09609-5>
- Cicchello, A. F., Marrazza, F., & Perdichizzi, S. (2023). Non-financial disclosure regulation and environmental, social, and governance (ESG) performance: The case of EU and US firms. *Corporate Social Responsibility and Environmental Management*, 30(3), 1121–1128. <https://doi.org/10.1002/csr.2408>
- CMF. (2020). CMF presents its Strategy for Addressing Climate Change in financial markets. Retrieved November 28, 2023, from <https://www.cmfchile.cl/portal/principal/613/w3-article-29874.html>
- CMF. (2021). Norma de Carácter General N°461 (General Rule No. 461). Retrieved November 28, 2023, from https://www.cmfchile.cl/normativa/ncg_461_2021.pdf
- CMF. (2022). CMF issues regulation incorporating sustainability and corporate governance requirements in Annual Reports. Retrieved November 28, 2023, from https://www.cmfchile.cl/portal/principal/613/articles-49809_recurso_1.pdf
- Cohen, J. R., & Simnett, R. (2015). CSR and Assurance Services: A Research Agenda. *AUDITING: A Journal of Practice & Theory*, 34(1), 59–74. <https://doi.org/10.2308/ajpt-50876>
- Cordazzo, M., Bini, L., & Marzo, G. (2020). Does the EU Directive on non-financial information influence the value relevance of ESG disclosure? Italian evidence. *Business Strategy and the Environment*, 29(8), 3470–3483. <https://doi.org/10.1002/bse.2589>
- Cousins, P., Dutordoir, M., Lawson, B., & Neto, J. Q. F. (2020). Shareholder Wealth Effects of Modern Slavery Regulation. *Management Science*, 66(11), 5265–5289. <https://doi.org/10.1287/mnsc.2019.3456>
- CSA. (2004). National Instrument 51-102 Continuous Disclosure Obligations. Retrieved November 26, 2023, from https://www.osc.ca/sites/default/files/pdfs/irps/rule_20040402_51-102-cont-disc-ob.pdf
- CSA. (2021). Proposed National Instrument 51-107: Disclosure of Climate-related Matters. Retrieved November 26, 2023, from https://www.osc.ca/sites/default/files/2021-10/csa_20211018_51-107_disclosure-update.pdf
- CSR Europe and GRI. (2017). Member State Implementation of Directive 2014/95/EU: A comprehensive overview of how Member States are implementing the EU Directive on Non-financial and Diversity Information. Retrieved September 20, 2023, from <https://www.accountancyeurope.eu/wp-content/uploads/NFR-Publication-3-May-revision.pdf>
- CSRC. (2021). Guidelines No. 2 on Contents and Formats of Information Disclosure by Companies Offering Securities to the Public - Contents and Formats of Annual Reports (Revised in 2021). Retrieved January 6, 2024, from <https://neris.csrc.gov.cn/falvfagui/rdqs/Header/mainbody?navbarId=3&secFutrLawId=3a5979eea68342819a3d9f5aad0af4d8&body=>
- Cuomo, F., Gaia, S., Girardone, C., & Piserà, S. (2022). The effects of the EU non-financial reporting directive on corporate social responsibility. *The European Journal of Finance*, 1–27. <https://doi.org/10.1080/1351847X.2022.2113812>
- CVM. (2021). Resolução CVM N° 59 (RCVM 59). Retrieved November 27, 2023, from <https://conteudo.cvm.gov.br/export/sites/cvm/legislacao/resolucoes/anexos/001/resol059consolid.pdf>
- CVM. (2023). Resolução CVM n° 193 (RCVM 193). Retrieved November 27, 2023, from <https://www.gov.br/cvm/en/foreign-investors/regulation-files/ResolutionCVM193.pdf>
- Dhaliwal, D. S., Li, O. Z., Tsang, A., & Yang, Y. G. (2011). Voluntary Non-financial Disclosure and the Cost of Equity Capital: The Initiation of Corporate Social Responsibility Reporting. *The Accounting Review*, 86(1), 59–100. <https://doi.org/10.2308/accr.00000005>
- Dharmapala, D., & Khanna, V. (2018). The impact of mandated corporate social responsibility: Evidence from India's Companies Act of 2013.

- International Review of Law and Economics*, 56, 92–104. <https://doi.org/10.1016/j.irle.2018.09.001>
- Doshi, A. R., Dowell, G. W. S., & Toffel, M. W. (2013). How firms respond to mandatory information disclosure. *Strategic Management Journal*, 34(10), 1209–1231. <https://doi.org/10.1002/smj.2055>
- Downar, B., Ernstberger, J., Reichelstein, S., Schwenen, S., & Zaklan, A. (2021). The impact of carbon disclosure mandates on emissions and financial operating performance. *Review of Accounting Studies*, 26(3), 1137–1175. <https://doi.org/10.1007/s11142-021-09611-x>
- Elayan, F. A., Brown, K., Li, J., & Chen, Y. (2021). The Market Response to Mandatory Conflict Mineral Disclosures. *Journal of Business Ethics*, 169(1), 13–42. <https://doi.org/10.1007/s10551-019-04283-9>
- Eliwa, Y., Aboud, A., & Saleh, A. (2021). ESG practices and the cost of debt: Evidence from EU countries. *Critical Perspectives on Accounting*, 79, 102097. <https://doi.org/10.1016/j.cpa.2019.102097>
- Engler, H. (2023). Companies need to integrate climate reporting across functions to comply with California's new law. Retrieved November 15, 2023, from <https://www.thomsonreuters.com/en-us/po sts/esg/california-climate-reporting-law/>
- EPRS. (2021). Implementation Appraisal: Non-financial Reporting Directive. Retrieved August 21, 2023, from [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/654213/EPRS_BRI\(2021\)654213_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/654213/EPRS_BRI(2021)654213_EN.pdf)
- Ernstberger, J., Maniora, J., & Schramm, F. (2021). *How to Regulate Corporate Social Responsibility Reporting - International Evidence on the Impact of Different Regulatory Approaches* [Research paper]. Technical University of Munich. Retrieved October 12, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3850179
- European Commission. (2014). Disclosure of non-financial and diversity information by large companies and groups - Frequently asked questions. Retrieved October 4, 2023, from https://ec.europa.eu/commission/presscorner/detail/en/MEMO_14_301
- European Commission. (2018a). Action Plan: Financing Sustainable Growth. Retrieved November 24, 2023, from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0097>
- European Commission. (2018b). Renewed sustainable finance strategy and implementation of the action plan on financing sustainable growth. Retrieved November 23, 2023, from https://finance.ec.europa.eu/publications/renewed-sustainable-finance-strategy-and-implementation-action-plan-financing-sustainable-growth_en#strategy
- European Commission. (2018c). Sustainable finance: Commission's Action Plan for a greener and cleaner economy. Retrieved September 19, 2023, from https://ec.europa.eu/commission/presscorner/detail/en/IP_18_1404
- European Commission. (2021). State and trends of ESG disclosure policy measures across IPSF jurisdictions, Brazil, and the US. Retrieved December 2, 2023, from https://finance.ec.europa.eu/system/files/2021-11/211104-ipsf-esg-disclosure-report_en.pdf
- European Commission. (2022a). Annual Report on European SMEs 2021/2022: SMEs and environmental sustainability background document. Retrieved December 12, 2023, from <https://op.europa.eu/en/publication-detail/-/publication/c45665ad-fd9a-11ec-b94a-01aa75ed71a1/language-en>
- European Commission. (2022b). Just and sustainable economy: Commission lays down rules for companies to respect human rights and environment in global value chains. Retrieved January 4, 2024, from https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1145
- European Commission. (2023a). Annex I - European Sustainability Reporting Standards (ESRS). Retrieved November 22, 2023, from https://ec.europa.eu/finance/docs/level-2-measures/csrd-delegated-act-2023-5303-annex-1_en.pdf
- European Commission. (2023b). Annual Report on European SMEs 2022/2023: SME Performance Review 2022/2023. Retrieved January 25, 2024, from https://single-market-economy.ec.europa.eu/system/files/2023-08/Annual%20Report%20on%20European%20SMEs%202023_FINAL.pdf
- European Commission. (2023c). Commission Delegated Regulation (EU) .../... of 31.7.2023 supplementing Directive 2013/34/EU of the European Parliament and of the Council as regards sustainability reporting standards (ESRS). Retrieved December 2, 2023, from https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13765-European-sustainability-reporting-standards-first-set_en
- European Commission. (2023d). Corporate sustainability reporting. Retrieved July 4, 2023, from https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en
- European Commission. (2023e). Factsheet: How does the EU Taxonomy fit within the sustainable finance framework? Retrieved August 1, 2023, from https://finance.ec.europa.eu/system/files/2021-04/sustainable-finance-taxonomy-factsheet_en.pdf
- European Commission. (2023f). Q&A on the Adoption of European Sustainability Reporting Standards. Retrieved November 23, 2023, from https://ec.europa.eu/commission/presscorner/detail/en/qanda_23_4043
- European Commission. (2023g). Sustainability-related disclosure in the financial services sector. Retrieved August 22, 2023, from https://finance.ec.europa.eu/sustainable-finance/disclosures/sustainability-related-disclosure-financial-services-sector_en
- European Commission. (2024). SME definition. Retrieved January 25, 2024, from https://single-market-economy.ec.europa.eu/smes/sme-definition_en
- European Parliament. (2023). Corporate Sustainability Reporting Directive (CSRD). Retrieved September 28, 2023, from <https://www.europarl.europa.eu/legislative-train/theme-a-european-green-deal/file-review-of-the-non-financial-reporting-directive>
- European Parliament and Council. (2003). Directive 2003/51/EC of the European Parliament and of the Council. Retrieved November 23, 2023, from <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:178:0016:0022:en:PDF>
- European Parliament and Council. (2013). Directive 2013/34/EU of the European Parliament and of the Council. Retrieved October 19, 2023, from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2013:182:FULL>
- European Parliament and Council. (2014). Directive 2014/95/EU of the European Parliament and of the Council (NFRD). Retrieved July 18, 2023, from <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32014L0095>
- European Parliament and Council. (2019). Regulation (EU) 2019/2088 of the European Parliament and of the Council (SFDR). Retrieved November 23, 2023, from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R2088>
- European Parliament and Council. (2020a). Assessing environmentally sustainable investments: Summary of the EU Taxonomy. Retrieved August 21, 2023, from <https://eur-lex.europa.eu/EN/legal-content/summary/assessing-environmentally-sustainable-investments.html>
- European Parliament and Council. (2020b). Regulation (EU) 2020/852 of the European Parliament and of the Council (EU Taxonomy). Retrieved November 23, 2023, from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32020R0852>
- European Parliament and Council. (2021). Commission Delegated Regulation (EU) 2021/2178. Retrieved August 22, 2023, from <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32021R2178#d1e363-9-1>
- European Parliament and Council. (2022). Directive (EU) 2022/2464 of the European Parliament and of the Council (CSRD). Retrieved September 28, 2023, from <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022L2464>
- European Union. (2016). Paris Agreement. Retrieved November 24, 2023, from [https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:22016A1019\(01\)](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:22016A1019(01))
- EY. (2023). Technical Line: How the climate-related disclosures under the SEC proposal, the ESRS and the ISSB standards compare. Retrieved December 2, 2023, from https://www.ey.com/en_us

- /assurance/accountinglink/technical-line---how-the-climate-related-disclosure-proposals-fr
- Fiechter, P., Hitz, J.-M., & Lehmann, N. (2022). Real Effects of a Widespread CSR Reporting Mandate: Evidence from the European Union's CSR Directive. *Journal of Accounting Research*, 60(4), 1499–1549. <https://doi.org/10.1111/1475-679X.12424>
- Flammer, C. (2015). Does Corporate Social Responsibility Lead to Superior Financial Performance? A Regression Discontinuity Approach. *Management Science*, 61(11), 2549–2568. <https://doi.org/10.1287/mnsc.2014.2038>
- FSA. (2023). Amendments on the Disclosure of Corporate Information. Retrieved November 17, 2023, from <https://www.fsa.go.jp/news/r4/sonota/20230131/20230131.html>
- GHG Protocol. (2023). Calculation Tools FAQ. Retrieved September 19, 2023, from <https://ghgprotocol.org/calculation-tools-faq>
- Government of Buenos Aires. (2007). Ley 2594. Retrieved January 5, 2024, from <https://boletinoficial.buenosaires.gob.ar/normativaba/norma/302800>
- Government of Canada. (2023). About the Greenhouse Gas Reporting Program. Retrieved November 26, 2023, from <https://www.canada.ca/en/environment-climate-change/services/climate-change/greenhouse-gas-emissions/facility-reporting/about.html>
- Government of France. (1977). La loi n°77-769 du 12 juillet 1977 (Bilan Social). Retrieved November 19, 2023, from <https://francearchives.gouv.fr/findingaid/018f07db1a7ea5752dca0e22d60cd7f4de254899>
- Government of France. (2001). LOI n° 2001-420 du 15 mai 2001 relative aux nouvelles régulations économiques (NRE). Retrieved November 18, 2023, from <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000223114>
- Government of France. (2009). LOI n° 2009-967 du 3 août 2009 de programmation relative à la mise en œuvre du Grenelle de l'environnement (Grenelle 1). Retrieved November 19, 2023, from <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT0000020949548/>
- Government of France. (2010). LOI n° 2010-788 du 12 juillet 2010 portant engagement national pour l'environnement (Grenelle 2, Article 225). Retrieved November 19, 2023, from https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000022471678
- Government of France. (2017a). LOI n° 2017-399 du 27 mars 2017 relative au devoir de vigilance des sociétés mères et des entreprises donneuses d'ordre (Devoir de Vigilance). Retrieved November 19, 2023, from <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000034290626/>
- Government of France. (2017b). Ordonnance n° 2017-1180 du 19 juillet 2017 relative à la publication d'informations non financières par certaines grandes entreprises et certains groupes d'entreprises. Retrieved November 20, 2023, from <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000035250851>
- Government of Japan. (1979). Act on the Rational Use of Energy. Retrieved November 18, 2023, from https://www.japaneselawtranslation.go.jp/en/laws/view/71#je_ch3sc1at11
- Gramlich, J., & Huang, L. (2017). *The Effect of Mandated CSR Disclosure on the Pollution Levels of Publicly-Traded Chinese Firms* [Research paper]. Washington State University. <https://doi.org/10.2139/ssrn.2906533>
- Greene, W. H. (2012). *Econometric Analysis* (7th edition). Pearson.
- Grewal, J., Riedl, E. J., & Serafeim, G. (2019). Market Reaction to Mandatory Nonfinancial Disclosure. *Management Science*, 65(7), 3061–3084. <https://doi.org/10.1287/mnsc.2018.3099>
- GSIA. (2023). Global Sustainable Investment Review 2022. Retrieved January 3, 2023, from <https://www.gsi-alliance.org/wp-content/uploads/2023/12/GSIA-Report-2022.pdf>
- Gulenko, M. (2018). Mandatory CSR reporting—literature review and future developments in Germany. *Nachhaltigkeits Management Forum | Sustainability Management Forum*, 26(1-4), 3–17. <https://doi.org/10.1007/s00550-018-0476-9>
- Gulenko, M. (2023). *Three Essays on Corporate Social Responsibility Activities and Reporting* [Research paper]. Paderborn University. Retrieved June 10, 2023, from <https://d-nb.info/1297297709/34#page=16>
- Haji, A. A., Coram, P., & Troshani, I. (2023). Consequences of CSR reporting regulations worldwide: a review and research agenda. *Accounting, Auditing & Accountability Journal*, 36(1), 177–208. <https://doi.org/10.1108/AAAJ-05-2020-4571>
- Healy, P. M., & Serafeim, G. (2020). Voluntary, Self-Regulatory, and Mandatory Disclosure of Oil and Gas Company Payments to Foreign Governments. *Accounting Horizons*, 34(1), 111–129. <https://doi.org/10.2308/horizons-17-133>
- HKEX. (2019). Hong Kong Main Board Listing Rules Appendix 27: Environmental, Social and Governance Reporting Guide (ESG Reporting Guide). Retrieved December 1, 2023, from https://en-rules.hkex.com.hk/sites/default/files/net_file_store/HKEX4476_3841_VER18584.pdf
- HKEX. (2021). FAQ Series 18. Retrieved December 11, 2023, from https://en-rules.hkex.com.hk/sites/default/files/net_file_store/HKEX_FAQ_18.pdf
- Ho, V. H. (2017). "Comply or explain" and the future of non-financial reporting. *Lewis & Clark Law Review*, 21(2), 317–355. Retrieved November 30, 2023, from <https://heinonline.org/HOL/LandingPage?handle=hein.journals/lewclr21&div=16&id=&page=>
- Hofstede, G. (1983). The Cultural Relativity of Organizational Practices and Theories. *Journal of International Business Studies*, (14), 75–89. <https://doi.org/10.1057/palgrave.jibs.8490867>
- Hummel, K., & Rötzel, P. (2019). Mandating the Sustainability Disclosure in Annual Reports—Evidence from the United Kingdom. *Schmalenbach Business Review*, 71(2), 205–247. <https://doi.org/10.1007/s41464-019-00069-8>
- IFRS. (2023a). About the International Sustainability Standards Board. Retrieved December 1, 2023, from <https://www.ifrs.org/groups/international-sustainability-standards-board/>
- IFRS. (2023b). ISSB issues inaugural global sustainability disclosure standards. Retrieved December 1, 2023, from <https://www.ifrs.org/news-and-events/news/2023/06/issb-issues-ifrs-s1-ifrs-s2/>
- Institute of Directors Southern Africa. (2016). King IV Report on Corporate Governance for South Africa 2016. Retrieved November 25, 2023, from https://cdn.ymaws.com/www.iodsa.co.za/resource/collection/684B68A7-B768-465C-8214-E3A007F15A5A/IoDSA_King_IV_Report_-_WebVersion.pdf
- Ioannou, I., & Serafeim, G. (2017). *The Consequences of Mandatory Corporate Sustainability Reporting* [Research paper]. London Business School. Retrieved August 3, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1799589
- Jackson, G., Bartosch, J., Avetisyan, E., Kinderman, D., & Knudsen, J. S. (2020). Mandatory Non-financial Disclosure and Its Influence on CSR: An International Comparison. *Journal of Business Ethics*, 162(2), 323–342. <https://doi.org/10.1007/s10551-019-04200-0>
- Japan Exchange Group. (2021). Publication of Revised Japan's Corporate Governance Code. Retrieved November 18, 2023, from <https://www.jpx.co.jp/english/news/1020/20210611-01.html>
- Japan Ministry of the Environment. (2006). Act on Promotion of Global Warming Countermeasures (Act No. 117 of 1998). Retrieved January 7, 2024, from https://www.japaneselawtranslation.go.jp/en/laws/view/3796#je_ch4at4
- Japan Ministry of the Environment. (2021). Introduction to Japan's Mandatory GHG Accounting and Reporting System. Retrieved November 18, 2023, from https://jprsi.go.jp/files/event_material/d1_3_03_mr_uchida.pdf
- Jeffery, C., Tenwick, J., & Bicciolo, G. (2017). Comparing the implementation of the EU Non-Financial Reporting Directive in the UK, Germany, France and Italy (F. Gregot, Ed.). Retrieved November 2, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3083368
- Jouvenot, V., & Krueger, P. (2019). *Mandatory Corporate Carbon Disclosure: Evidence from a Natural Experiment* [Research paper]. University of Geneva. <https://doi.org/10.2139/ssrn.3434490>
- JSE. (2022). JSE Limited Listings Requirements. Retrieved November 25, 2023, from <https://www.jse.co.za/sites/default/files/media/documents/jse-listings-requirements-incl-bulletin-1-2023/JSE%20Listings%20Requirements%2C%20incl%20Bulletin%201%20f%202023.pdf>

- Kinderman, D. (2020). The challenges of upward regulatory harmonization: The case of sustainability reporting in the European Union. *Regulation & Governance*, 14(4), 674–697. <https://doi.org/10.1111/rego.12240>
- Krueger, P. (2015). Climate change and firm valuation: Evidence from a quasi-natural experiment. *Swiss Finance Institute Research Paper*, 1–58. Retrieved October 13, 2023, from <https://ou.edu/content/dam/price/Finance/energyfinanceconference/Climate%20Change%20and%20Firm%20Valuation.pdf>
- Krueger, P., Sautner, Z., Tang, D. Y., & Zhong, R. (2023). *The Effects of Mandatory ESG disclosure around the world* [Working paper]. University of Geneva. <http://ssrn.com/abstractid=3832745>
- Lavigne, B., Motta, C., Monteiro, D. R., Montes, L., & Bezerra, L. G. (2021). Brazilian Securities Commission Establishes ESG Information Disclosure Criteria for Listed Companies. Retrieved November 1, 2023, from <https://www.lexology.com/library/detail.aspx?g=1c5f6f56-62a1-40d4-9136-8639f2dde0da>
- Lawless, K. P., Ushijima, K., & Hara, K. (2023). What's next for Japanese sustainability disclosure standards. Retrieved November 18, 2023, from https://www.ey.com/en_jp/sustainability/whats-next-for-japanese-sustainability-disclosure-standards
- Leuz, C. (2010). Different Approaches to Corporate Reporting Regulation: How Jurisdictions Differ and Why. *Accounting and Business Research*, 40(3), 229–256. Retrieved November 28, 2023, from https://www.ecgi.global/sites/default/files/working_papers/documents/SSRN-id1581472.pdf
- Leuz, C., & Wysocki, P. (2016). The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research. *Journal of Accounting Research*, 54(2), 525–622. <https://doi.org/10.1111/1475-679X.12115>
- 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 of Cleaner Production*, 122, 186–200. <https://doi.org/10.1016/j.jclepro.2016.02.060>
- LSEG. (2022). Environmental, social and governance scores from LSEG. Retrieved December 8, 2023, from https://www.lseg.com/content/dam/data-analytics/en_us/documents/methodology/lseg-esg-scores-methodology.pdf
- LSEG. (2023a). The Refinitiv Acquisition - LSEG is Open. Retrieved December 8, 2023, from <https://www.lseg.com/en/media-centre/refinitiv-acquisition>
- LSEG. (2023b). Worldscope Fundamentals. Retrieved January 7, 2024, from <https://www.lseg.com/en/data-analytics/financial-data/company-data/fundamentals-data/worldscope-fundamentals>
- Lu, T., Sivaramakrishnan, K., Wang, Y., & Yu, L. (2021). The Real Effects of Mandatory Corporate Social Responsibility Reporting in China. *Production and Operations Management*, 30(5), 1493–1516. <http://doi.org/10.1111/poms.13334>
- Manchiraju, H., & Rajgopal, S. (2017). Does Corporate Social Responsibility (CSR) Create Shareholder Value? Evidence from the Indian Companies Act 2013. *Journal of Accounting Research*, 55(5), 1257–1300. <https://doi.org/10.1111/1475-679X.12174>
- Matisoff, D. C. (2013). Different rays of sunlight: Understanding information disclosure and carbon transparency. *Energy Policy*, 55, 579–592. <https://doi.org/10.1016/j.enpol.2012.12.049>
- MCA. (2013). The Companies Act, 2013. Retrieved September 13, 2023, from <https://www.mca.gov.in/Ministry/pdf/CompaniesAct2013.pdf>
- Mcgowan, J. (2023). The SEC May Be Overstepping Its Authority In ESG/Climate Related Disclosure Standards. Retrieved November 13, 2023, from <https://www.forbes.com/sites/jonmcgowan/2023/08/30/the-sec-may-be-overstepping-its-authority-in-esg-climate-related-disclosure-standards/?sh=2c946d301a2f>
- MEE. (2021a). Administrative Measures for Legal Disclosure of Corporate Environmental Information. Retrieved November 11, 2023, from https://www.mee.gov.cn/xxgk2018/xxgk/xxgk02/202112/t20211221_964837.html?mc_cid=45e6a7ad33&mc_eid=627c47469b
- MEE. (2021b). Notice by the Ministry of Ecology and Environment of Issuing the Plan for the Reform of the Legal Disclosure System of Environmental Information. Retrieved November 12, 2023, from <https://lawinfochina.com/display.aspx?id=36050&lib=law>
- Ministère des Affaires Étrangères France. (2012). The French legislation on extra-financial reporting. Retrieved August 22, 2023, from https://www.diplomatie.gouv.fr/IMG/pdf/Mandatory_reporting_build_on_consensus_in_France.pdf
- Mion, G., & Adaii, C. R. L. (2020). The Effect of Mandatory Publication of Nonfinancial Disclosure in Europe on Sustainability Reporting Quality: First Insights about Italian and German Companies. *Studies in Managerial and Financial Accounting*, 34, 55–80. <https://doi.org/10.1108/S1479-351220200000034005>
- Mittelbach-Hörmanseder, S., Hummel, K., & Rammerstorfer, M. (2021). The information content of corporate social responsibility disclosure in Europe: an institutional perspective. *European Accounting Review*, 30(2), 309–348. <https://doi.org/10.1080/09638180.2020.1763818>
- Murray, S. (2021). Measuring what matters: the scramble to set standards for sustainable business. Retrieved December 5, 2023, from <http://www.ft.com/content/92915630-c110-4364-86ee-0f6f018c6a90>
- Nasdaq. (2023). Nasdaq's Board Diversity Rule. Retrieved November 13, 2023, from <https://listingcenter.nasdaq.com/assets/Board%20Diversity%20Disclosure%20Five%20Things.pdf>
- OECD. (2023). Sustainability Policies and Practices for Corporate Governance in Asia: Sustainability disclosure. Retrieved January 6, 2024, from <https://www.oecd-ilibrary.org/sites/cb598d94-en/index.html?itemId=/content/component/cb598d94-en>
- Ottenstein, P., Erben, S., Jost, S., Weuster, C. W., & Zülch, H. (2022). From voluntarism to regulation: effects of Directive 2014/95/EU on sustainability reporting in the EU. *Journal of Applied Accounting Research*, 23(1), 55–98. <https://doi.org/10.1108/JAAR-03-2021-0075>
- Parliament of the United Kingdom. (2006a). Companies Act 2006 (Part 15). Retrieved September 16, 2023, from <https://www.legislation.gov.uk/ukpga/2006/46/part/15/data.pdf>
- Parliament of the United Kingdom. (2006b). Companies Act 2006 (Section 417). Retrieved September 23, 2023, from <https://www.legislation.gov.uk/ukpga/2006/46/section/417/2007-10-01?view=plain>
- Parliament of the United Kingdom. (2013). The Companies Act 2006 (Strategic Report and Directors' Report) Regulations 2013. Retrieved September 25, 2023, from https://www.legislation.gov.uk/uksi/2013/1970/pdfs/uksi_20131970_en.pdf
- Parliament of the United Kingdom. (2016). The Companies, Partnerships and Groups (Accounts and Non-Financial Reporting) Regulations 2016. Retrieved November 20, 2023, from <https://www.legislation.gov.uk/uksi/2016/1245/made>
- Parliament of the United Kingdom. (2022). The Companies (Strategic Report) (Climate-related Financial Disclosure) Regulations 2022. Retrieved November 21, 2023, from <https://www.legislation.gov.uk/uksi/2022/31/note/made>
- PwC India. (2021). Mainstreaming ESG via Business Responsibility and Sustainability Reporting. Retrieved November 16, 2023, from <https://www.pwc.in/assets/pdfs/consulting/esg/business-responsibility-and-sustainability-report.pdf>
- R Documentation. (2024). Winsorize data. Retrieved January 15, 2024, from <https://search.r-project.org/CRAN/refmans/datawizard/html/winsorize.html>
- Raimo, N., Caragnano, A., Zito, M., Vitolla, F., & Mariani, M. (2021). Extending the benefits of ESG disclosure: The effect on the cost of debt financing. *Corporate Social Responsibility and Environmental Management*, 28(4), 1412–1421. <https://doi.org/10.1002/csr.2134>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. Retrieved November 3, 2023, from <https://academic.oup.com/biomet/article/70/1/41/240879>
- SASAC. (2008). Notification on the Issuance of the Guideline in Fulfilling Social Responsibility by Central Enterprises. Retrieved November 10, 2023, from <http://www.sasac.gov.cn/n2588030/n2588939/c4297490/content.html>

- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia and analgesia*, 126(5), 1763–1768.
- SEBI. (2012). Business Responsibility Reports (BRR). Retrieved August 17, 2023, from https://www.sebi.gov.in/sebi_data/attachdocs/1344915990072.pdf
- SEBI. (2021a). Business responsibility and sustainability reporting by listed entities. Retrieved August 17, 2023, from https://www.sebi.gov.in/legal/circulars/may-2021/business-responsibility-and-sustainability-reporting-by-listed-entities_50096.html
- SEBI. (2021b). Business responsibility and sustainability reporting format. Retrieved November 15, 2023, from https://www.sebi.gov.in/sebi_data/commondocs/may-2021/Business%20responsibility%20and%20sustainability%20reporting%20by%20listed%20entitiesAnnexure1_p.PDF
- SEBI. (2021c). Guidance note for business responsibility and sustainability reporting format. Retrieved November 16, 2023, from https://www.sebi.gov.in/sebi_data/commondocs/may-2021/Business%20responsibility%20and%20sustainability%20reporting%20by%20listed%20entitiesAnnexure2_p.PDF
- SEBI. (2023). BRSR Core - Framework for assurance and ESG disclosures for value chain. Retrieved August 17, 2023, from https://www.sebi.gov.in/legal/circulars/jul-2023/brsr-core-framework-for-assurance-and-esg-disclosures-for-value-chain_73854.html
- SEC. (2010). Commission Guidance Regarding Disclosure Related to Climate Change (Interpretation). Retrieved August 7, 2023, from <https://www.sec.gov/files/rules/interp/2010/33-9106.pdf>
- SEC. (2012a). Conflict Minerals. Retrieved November 13, 2023, from <https://www.sec.gov/files/rules/final/2012/34-67716.pdf>
- SEC. (2012b). Conflict Minerals Disclosure. Retrieved August 14, 2023, from <https://www.sec.gov/info/smallbus/secg/conflict-minerals-disclosure-small-entity-compliance-guide.htm>
- SEC. (2012c). Mine Safety Disclosure. Retrieved January 22, 2024, from <https://www.sec.gov/files/rules/final/2011/33-9286.pdf>
- SEC. (2017). Disclosing the Use of Conflict Minerals. Retrieved August 14, 2023, from <https://www.sec.gov/opa/Article/2012-2012-163htm---related-materials.html>
- SEC. (2020). Disclosure of Payments by Resource Extraction Issuers. Retrieved November 13, 2023, from <https://www.sec.gov/files/rules/final/2020/34-90679.pdf>
- SEC. (2021). Statement on Nasdaq's Diversity Proposals – A Positive First Step for Investors. Retrieved August 12, 2023, from <https://www.sec.gov/news/public-statement/statement-nasdaq-diversity-080621>
- SEC. (2022a). Fact Sheet - Enhancement and Standardization of Climate-Related Disclosures. Retrieved March 8, 2023, from <https://www.sec.gov/files/33-11042-fact-sheet.pdf>
- SEC. (2022b). The Enhancement and Standardization of Climate-Related Disclosures for Investors (Proposed Rule): SEC Proposed Climate Disclosure Rule. Retrieved November 13, 2023, from <https://www.sec.gov/files/rules/proposed/2022/33-11042.pdf>
- SEC. (2023). SEC Adopts Rules on Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure by Public Companies. Retrieved August 12, 2023, from <https://www.sec.gov/news/press-release/2023-139>
- She, G. (2022). The Real Effects of Mandatory Nonfinancial Disclosure: Evidence from Supply Chain Transparency. *The Accounting Review*, 97(5), 399–425. <https://doi.org/10.2308/TAR-2020-0178>
- SSE. (2008). Notice of Doing a Better Job for Disclosing 2008 Annual Reports. Retrieved November 6, 2023, from <http://english.sse.com.cn/news/newsrelease/c/4946972.shtml>
- Stolowy, H., & Paugam, L. (2018). The expansion of non-financial reporting: an exploratory study. *Accounting and Business Research*, 48(5), 525–548. <https://doi.org/10.1080/00014788.2018.1470141>
- Swift, C., Guide, V. D. R., & Muthulingam, S. (2019). Does supply chain visibility affect operating performance? Evidence from conflict minerals disclosures. *Journal of Operations Management*, 65(5), 406–429. <https://doi.org/10.1002/joom.1021>
- SZSE. (2008). Notice on the Preparation of 2008 Annual Reports of Listed Companies. Retrieved November 10, 2023, from https://www.szse.cn/disclosure/notice/general/t20081231_500181.html
- Tomar, S. (2023). Greenhouse Gas Disclosure and Emissions Benchmarking. *Journal of Accounting Research*, 61(2), 451–492. <https://doi.org/10.1111/1475-679X.12473>
- TSE. (2021). Japan's Corporate Governance Code. Retrieved November 18, 2023, from <https://www.jp.co.jp/english/news/1020/b5b4pj000046kxj-att/b5b4pj0000046l0c.pdf>
- UK Department for Business, Energy and Industrial Strategy. (2022). Mandatory climate-related financial disclosures by publicly quoted companies, large private companies and LLPs. Retrieved September 27, 2023, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1056085/mandatory-climate-related-financial-disclosures-publicly-quoted-private-cos-llps.pdf
- UK Government. (2023a). Gender pay gap reporting: guidance for employers. Retrieved September 27, 2023, from <https://www.gov.uk/government/publications/gender-pay-gap-reporting-guidance-for-employers>
- UK Government. (2023b). Smarter regulation non-financial reporting review: call for evidence. Retrieved September 26, 2023, from <https://www.gov.uk/government/calls-for-evidence/smarter-regulation-non-financial-reporting-review-call-for-evidence#fnref:1>
- UNEP. (2019). Background to sustainability reporting: Section A - Handbook. Retrieved November 27, 2023, from <https://wedocs.unep.org/bitstream/handle/20.500.11822/30663/CSRHdb.pdf?sequence=1&isAllowed=y>
- United Nations. (2015). Transforming our world: the 2030 Agenda for Sustainable Development. Retrieved November 24, 2023, from <http://sdgs.un.org/2030agenda>
- United Nations Global Compact. (2004). Who Cares Wins: Connecting Financial Markets to a Changing World. Retrieved February 1, 2024, from https://www.unepfi.org/fileadmin/events/2004/stocks/who_cares_wins_global_compact_2004.pdf
- US Congress. (2010). Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank Act). Retrieved November 13, 2023, from <https://www.govinfo.gov/content/pkg/PLAW-111publ203/pdf/PLAW-111publ203.pdf>
- US EPA. (2009). EPA Finalizes the Nation's First Greenhouse Gas Reporting System/Monitoring to begin in 2010. Retrieved August 3, 2023, from https://www.epa.gov/archive/epapages/newsroom_archive/newsreleases/194e412153fcfea8525763900530d75.html
- US EPA. (2023a). Fact Sheet: Greenhouse Gas Reporting Program (GHGRP) Implementation. Retrieved August 15, 2023, from <https://www.epa.gov/system/files/documents/2023-08/ghgrp-overview-fact-sheet.pdf>
- US EPA. (2023b). Greenhouse Gas Reporting Program (GHGRP) | US EPA. Retrieved August 3, 2023, from <https://www.epa.gov/ghgreporting>
- Veltri, S., de Luca, F., & Phan, H.-T.-P. (2020). Do investors value companies' mandatory nonfinancial risk disclosure? An empirical analysis of the Italian context after the EU Directive. *Business Strategy and the Environment*, 29(6), 2226–2237. <https://doi.org/10.1002/bse.2497>
- Xu, H., Xu, X., & Yu, J. (2021). The Impact of Mandatory CSR Disclosure on the Cost of Debt Financing: Evidence from China. *Emerging Markets Finance and Trade*, 57(8), 2191–2205. <https://doi.org/10.1080/1540496X.2019.1657401>



The Impact of Female Board Members on ESG Performance: An Empirical Analysis

Finn Matthes Gooßen

University of Hamburg

Abstract

This study investigates the impact of female board representation on ESG performance within the German two-tier corporate governance system. Using OLS regression analysis on a sample of 157 DAX, MDAX, and SDAX companies over a two-year period, the findings reveal a positive and significant relationship between the presence of women on both management and supervisory boards and improved ESG performance, regardless of whether measured by percentage or absolute number. Contrary to the critical mass theory, even the presence of a single woman on a board was found to significantly enhance ESG outcomes. No statistically significant effect was observed for female CEOs, although this is likely to be attributed to the low number of female CEOs in the sample. The study highlights both the ongoing underrepresentation of women on boards and the limited scope of current gender quota regulations (FüPoG I & II), suggesting the need for stronger legislative measures to support gender diversity as a driver of corporate sustainability.

Keywords: board gender diversity; board structure; critical mass; corporate governance; ESG performance

1. Introduction

In times of climate change and social inequality, sustainability is becoming an increasingly important issue for companies. More and more companies are pursuing goals and implementing initiatives to increase sustainability to meet consumer expectations and regulatory requirements. Not only is it a moral imperative, but it also has evolved into a strategic priority across industries.¹ Sustainability is of great significance to both shareholders and stakeholders. Recent data shows that 85 per cent of investors include sustainability in their investment decisions.² Furthermore, according to legitimacy theory, companies engage in sustainability activities to align with expectations of stakeholders and society in general, to secure their social approval and therefore their

continued existence.³ For this reason, companies should focus on sustainability and look for opportunities and ways to improve their ESG performance, both as a strategic necessity and a driver of competitive advantage.

Women on management and supervisory boards of German DAX, MDAX and SDAX companies are severely underrepresented,⁴ although research shows that gender diversity and the presence of women on boards has a positive impact on a wide range of financial and non-financial metrics.⁵ These findings have led German legislators to pass two laws aimed at increasing the proportion of women in management positions of listed companies with codetermination; FüPoG I which establishes a fixed quota of 30 per cent for both genders on the supervisory board⁶ and FüPoG II which requires at least one member of each gender for management boards with more than three members.⁷ In this context, the question arises as to what extent an increased pro-

I would like to express my gratitude towards Professor Nicole Ratzinger-Sakel for her support and guidance along my academic journey and for giving me the opportunity to write this master's thesis. Additionally I would like to thank my supervisor Lars-Eric Brüske for his valuable ideas and feedback throughout this research.

¹ Cf. Farri et al. (2022), n.p. for these sentences.

² Cf. Gartner (Ed.) (2021), n.p.

³ Cf. Crossley et al. (2021), p. 3754.

⁴ Cf. Frauen in die Aufsichtsräte e.V. (Ed.) (2024), p. 4–5.

⁵ Cf. Arayssi et al. (2016), p. 391–392.

⁶ Cf. section 96 Abs. 2 AktG.

⁷ Cf. section 76 Abs. 3a AktG.

portion of women impacts the ESG performance of the companies where the laws intended for that purpose are mainly applied. This study adds to the existing literature by examining the impact of women on ESG performance from four different perspectives, for the same sample of German DAX, MDAX and SDAX companies, 5 years after the introduction of FÜPoG I and shortly after FÜPoG II. Firstly, the influence of the proportion of women on the management board is analysed. Secondly, the study addresses the question of whether a certain threshold of women, which is referred to as a 'critical mass', must be reached to have a positive effect. Thirdly, it examines whether female CEOs positively influence a company's sustainability. Finally, the impact of women on the supervisory board on ESG performance is analysed.

The remainder of this thesis is structured as follows. In section 2 the conceptual background is described, defining relevant terms and describing the regulatory environment in Germany with a particular focus on women on the management and supervisory board. In section 3 the existing literature on the topic is reviewed and from it, four hypotheses are developed. Section 4 describes the research design and statistical model used, to test these hypotheses. The descriptive statistics and results of the regressions are presented in section 5. The results are then put into the context of the previous research and limitations are given. These limitations are addressed by conducting robustness tests which are outlined and discussed in section 6. Additionally further analysis is done, regarding the circumstances that regulate the impact women on boards have on ESG performance. Finally, conclusions are drawn, remaining limitations that could not be addressed by the robustness tests given, and recommendations for actions as well as for future research made.

2. Conceptual and Regulatory Background

2.1. Basic Definitions

Although there is no standardised and universally agreed definition of sustainability, the definition of the Brundtland Commission of the United Nations (1987) is widely accepted, which defines sustainability as „(...) meet(ing) the needs of the present without compromising the ability of future generations to meet their own needs“.⁸ The sustainability of a company can be viewed in terms of Elkington's (2004) triple bottom line concept, measuring a business's success in three key areas: people, planet and profit, also known as the 3 Ps.⁹ This implies that economic success can no longer be measured solely in terms of profits, but rather represents one of three dimensions, alongside social and environmental factors. Taking up the triple bottom line concept companies measure sustainability using environmental, social and governance (ESG) measures.¹⁰ In light of this, the terms 'ESG

performance' and 'sustainability performance' are used interchangeably in the context of this study. The actions of companies can have positive or adverse effects on environmental, social and economic systems. ESG performance, the main focus of this study, can be defined as the extent to which a company's actions affect the long-term viability and health of these systems positively or negatively.¹¹

Corporate social responsibility (CSR) is a closely related concept. The international standard ISO 26000 defines CSR as the responsibility of organizations for the effects their actions have on society and the environment through transparent and ethical conduct.¹² Differentiating CSR from sustainability, the definition of the ISO standard shows that CSR is rather to be understood as a framework of actions to achieve the normative goal of sustainability. ESG performance can therefore be seen as a link between the two concepts: it indicates the extent to which a company fulfils its sustainability obligations.

2.2. Regulatory

To increase the proportion of women in management positions, the first Leadership Positions Act (FÜPoG I) was passed in 2015 and came into force on 1 January 2016. It was intended to enable the equal participation of women in economic life.¹³ To this end, a gender quota of at least 30 per cent was set for the supervisory board. This was implemented by requiring new appointments to be made with the underrepresented gender. However, the regulation only applies to listed companies that are also subject to codetermination as they have more than 2,000 employees.¹⁴ Companies that fulfil at least one of these criteria must also set targets for the proportion of women on the management board and upper management levels.¹⁵ A study from 2020, 5 years after FÜPoG I came into force, shows that 115 of the 188 companies analysed, still have no women on their management boards. Of these companies, 75 have set a target figure of zero. Although the proportion of women has doubled since the introduction of the FÜPoG I, it is still only 10.7 per cent.¹⁶

With regard to increasing the participation of women on the management board, there was clearly still a need for further regulation. To this end, the Second Leadership Positions Act (FÜPoG II) was enacted, which came into force on 1 August 2022. Companies that fulfil the conditions of FÜPoG I and whose management board size exceeds three members must have at least one male and one female board member.¹⁷ The minimum participation requirement must be adhered to for new appointments: if the management board has more

¹¹ Cf. Zimek and Baumgartner (2017), p. 1–2 for these sentences.

¹² Cf. International Organization for Standardization (Ed.) (n.d.), n.p.

¹³ Cf. Deutscher Bundestag (Ed.) (2021), p. 1 for these sentences.

¹⁴ Cf. section 96 (2) sentence 1 AktG for these sentences.

¹⁵ Cf. section 76 (4) sentence 1 AktG.

¹⁶ Cf. Frauen in die Aufsichtsräte e.V. (Ed.) (2020), p. 3 for these sentences.

¹⁷ Cf. section 76 (3a) sentence 1 AktG.

⁸ United Nations (1987), n.p.

⁹ Cf. Elkington (2004), p. 2.

¹⁰ Cf. Grace and Gehman (2023), p. 151.

than three members and no woman is a member of the management board, only one woman can be effectively appointed as a member of the management board; the appointment of another male member would be null and void. An existing board composition does not have to be adjusted, mandates can remain in place until the end of the appointment.¹⁸

Furthermore, although the target figure can still be set as zero, the supervisory board then must give clear and comprehensible reasons and explain the considerations that led to this decision in the company's management report. The target figure for the management board is set by the supervisory board.¹⁹ Furthermore, the corporate governance statement must include whether the quota of women on the management board and supervisory board has been complied with, how both boards are composed and whether target figures have been met.²⁰ Violations of these reporting requirements can be penalised with a fine.²¹

At the end of 2022, 84 per cent of management board members in the DAX, MDAX and SDAX were male. A slightly positive trend can be observed compared to the previous year: in the period from 2021 to 2022, the percentage of female management board members rose from 14 to 16 per cent. Of these, the share of women in the DAX is the highest at 23 per cent and also the highest relative increase compared to the previous year (19 per cent). This is followed by the MDAX with 13 per cent (11 per cent in the previous year) and the SDAX with 12 per cent, where no increase compared to the previous year can be seen. Under the FÜPoG II, the question of whether at least one woman is represented on the management board is even more relevant. Here too, the DAX is far ahead with 93 per cent, with a large distance to the MDAX and SDAX where only 44 per cent and 42 per cent have at least one woman on the management board, respectively. In 2022, the year in which FÜPoG II came into force, the proportion across all three indices increased only slightly, from 53 per cent to 56 per cent. On supervisory boards where the FÜPoG I has been in place since 2015, the quota of female board members increased from 32 per cent in 2021 to 34 per cent within a year. Here the same pattern emerges: the DAX has the highest share at 37 per cent, but in contrast to the management board, the differences to the MDAX (33 per cent) and SDAX (31 per cent) are much smaller.²² As of 2024, 65 of the 104 listed companies that fall under the Codetermination Act are subject to the minimum participation requirement as they have more than three management board members. Of these 65 companies, three still do not have a woman on the management board.²³

There exist multiple models of corporate governance; the single-tiered model and two-tiered model being two of the

most widely used. While the one-tier system is the most widespread in Anglo-Saxon countries, the two-tier system is the most common in European jurisdictions, one of them being Germany.²⁴

In the German two-tier system, there are two boards: the supervisory board (Aufsichtsrat) and the management board (Vorstand). The two boards are separated, a member cannot belong to both boards at the same time.²⁵ Members of the management board are appointed by the supervisory board.²⁶ The members of the supervisory board are elected by the shareholders at the general meeting.²⁷ The management board must manage the company under its own responsibility.²⁸ The supervisory board's task is to supervise the management board²⁹; thus it fulfils a control function. Furthermore, the supervisory board must approve decisions that are of fundamental importance to the company.³⁰ For stock corporations with more than 500 employees, the One-Third Participation Act applies, which requires that one-third of the supervisory board must be made up of employee representatives³¹, in corporations with more than 2000 employees, representatives of workers must make up half of the supervisory board members.³²

3. Previous Literature and Hypothesis Development

The following section provides an overview of the existing research on women in management positions and their impact on ESG performance. To this end, theories are first presented that explain the potential mechanism that leads to women having a positive impact on sustainability. The literature on women on the management board, the supervisory board, in the position of CEO and on critical mass theory is then presented and hypotheses are derived from this.

First, according to principal-agent theory one party, the principal, delegates tasks to another party, the agent. In this configuration, a conflict of interest can arise, when the agent pursues their own goals and interests and acts opportunistically, contrary to the interests of the principal. Since the agent's actions are not fully observable by the principle, information asymmetry arises.³³

Secondly, stakeholder theory, which was first formulated by Hill and Jones (1992), states that companies should create value for all their stakeholders; these are, among others, employees, suppliers, customers and local communities. Stakeholders supply the firm with critical resources and in exchange can expect their needs to be satisfied by the firm,

²⁴ Cf. Ahmad and Omar (2016), p. 76–77 for these sentences.

²⁵ Cf. section 105 (1) AktG.

²⁶ Cf. section 84 (1) sentence 1 AktG.

²⁷ Cf. section 101 (1) sentence 1 AktG.

²⁸ Cf. section 76 (1) AktG.

²⁹ Cf. section 111 (1) AktG.

³⁰ Cf. section 111 (4) AktG.

³¹ Cf. section 1 (1) no. 1 in conjunction with section 4 (1) DrittelbG.

³² Cf. section 1 in conjunction with section (7) 1 MitbestG.

³³ Cf. Jensen and Meckling (1976), p. 308–309.

¹⁸ Cf. section 76 (3a) sentence 2 AktG for these sentences.

¹⁹ Cf. section 111 (5) AktG for these sentences.

²⁰ Cf. section 289f (2) no. 3, 4 HGB.

²¹ Cf. section 334 (1) no. 3a HGB.

²² Cf. BDO AG Wirtschaftsprüfungsgesellschaft and Kirchhoff Consult AG (Ed.) (2023), p. 6 for this paragraph.

²³ Cf. Frauen in die Aufsichtsräte e.V. (Ed.) (2024), p. 27 for these sentences.

thereby entering an implicit contract. An example the authors give is that the public provides firms with infrastructure through paying taxes and in exchange, they can expect that the firm does not reduce their quality of life through its actions.³⁴

Bringing together both theories, stakeholder-agency theory posits that managers are the only stakeholders who have a direct influence on the company's decisions. Therefore, they act as agents of the other stakeholder groups to enforce their interests.³⁵ If firms breach the implicit contracts they entered, they risk losing legitimacy. The risk of losing legitimacy and being unable to continue operation provides the firm with an incentive to act sustainably.³⁶

According to upper echelons theory, top managers' experiences and so-called managerial background characteristics have a partial influence on organizational outcomes. Upper echelons characteristics are both psychological base values and observable characteristics of the management personnel like age or education. While Hambrick and Mason (1984), who first formulated the upper echelons theory named managers' characteristics like age, or socioeconomic background as an influence on management outcomes; the directors' gender as a factor that determines the company's management outcomes was not examined in the original study.³⁷

From upper echelon theory it follows, that it is important to highlight the values and psychological attributes in which women differ from men to see what influence women have on the company. The underlying characteristics and moral attitudes of the members of the management board therefore have a significant impact on the extent to which the company acts sustainably and thus fulfils the needs of the other stakeholder groups. Adams and Funk (2012) find that women on boards differ in their values from women in the general population. While in the general population, men are more open to innovation, this relationship is reversed in boards.³⁸ This could drive change processes with regard to sustainable developments in the company.

Women are more benevolent than men; the well-being of others is important to them, security and tradition are less important to women than to men.³⁹ Being more benevolent, it is more likely that when there is a conflict between shareholders and stakeholders, women side with the stakeholders.⁴⁰ Furthermore, female directors are more sensitive to environmental concerns compared to male directors; gender-diverse boards report more transparently on ecological initiatives, are more involved with stakeholders and communicate more openly with them. Women play a key role in the implementation and success of the initiatives.⁴¹ Women are

more likely than men to have worked in non-profit organisations and therefore focus less on shareholders and more on community interests.⁴² They are more inclined to give something back to society, which benefits the company's reputation and corporate relations in the long term. This also benefits the company in economic terms.⁴³ Female directors tend to be more avoidant of litigation and reputation loss,⁴⁴ which might cause them to act in line with legislation concerning sustainability.

Isidro and Sobral (2015) find that female board presence is significantly and positively associated with ethical and social compliance, which is valued by investors and consequently increases firm value.⁴⁵ Kidwell et al. (1987) find that while male and female managers act the same when faced with an ethical decision, males were more inclined to conceal their errors, while females thought of that behaviour as unethical.⁴⁶ This tendency may lead females to foster and promote a culture of transparency and accountability. Based on these theories and gender differences, it can be theoretically explained in what way women on boards have a positive impact in terms of ESG performance.

Turning to empirical research on the influence of women on ESG performance, studies that analyse the impact of women in top management positions on corporate sustainability can be divided into four basic categories: sustainable activity, sustainable performance, sustainable disclosure and finally the impact of sustainability performance on firm performance and value.⁴⁷ These domains build on each other logically; top management influences sustainability activities, which in turn have an impact on sustainability performance. The accuracy and scope with which the company reports on activities and performance are measured by the sustainability disclosure.

Mattingly and Berman (2006) identify four classes of CSR, using factor analysis: institutional strengths and weaknesses and technical strengths and weaknesses. The institutional dimension describes measures towards stakeholders that have normative expectations of the company, such as local communities. The technical dimension, on the other hand, describes measures in relation to stakeholders with whom the company exchanges resources, for example, consumers or employees.⁴⁸ Building on this, Bear et al. (2010) analyse the relationship between board gender composition, CSR and reputation. They find that women on the board have a significant and positive influence on the institutional strength, but not on the technical strength. Women on the board also have a positive influence on firm reputation; this relationship is mediated by institutional strength, but not by technical strength.⁴⁹ Zhang (2012) finds that board diver-

³⁴ Cf. Hill and Jones (1992), p. 133 for this paragraph.

³⁵ Cf. Hill and Jones (1992), p. 134 for these sentences.

³⁶ Cf. Hrasaky (2011), p. 179 for these sentences.

³⁷ Cf. Hambrick and Mason (1984), p. 198 for this paragraph.

³⁸ Cf. Adams and Funk (2012), p. 228 for these sentences.

³⁹ Cf. Adams and Funk (2012), p. 226–227.

⁴⁰ Cf. Adams and Funk (2012), p. 231.

⁴¹ Cf. Haque and Jones (2020), p. 15 for these sentences.

⁴² Cf. Williams (2003), p. 2.

⁴³ Cf. Williams (2003), p. 8–9 for these sentences.

⁴⁴ Cf. Srinidhi et al. (2011), p. 1614.

⁴⁵ Cf. Isidro and Sobral (2015), p. 13.

⁴⁶ Cf. Kidwell et al. (1987), p. 490.

⁴⁷ Cf. Bannò et al. (2023), p. 191.

⁴⁸ Cf. Mattingly and Berman (2006), p. 34–37 for these sentences.

⁴⁹ Cf. Bear et al. (2010), p. 217 for these sentences.

sity has a positive and significant impact on technical and institutional strength ratings, for technical and institutional weaknesses no significant relationship can be found.⁵⁰ Harjoto et al. (2015) examine the effect of multiple diversity characteristics on ESG performance and find that tenure, age and gender diversity have the most positive impact. In contrast to tenure and age which only reduce CSR concerns, gender diversity additionally increases CSR strengths.⁵¹

Velte (2016b) was the first to examine the impact of the management board's gender diversity for the German and Austrian settings. The study's sample consisted of 1019 firm-year observations for the period 2011-2014. A significant and positive impact of female members on the management board on ESG performance was found. It also found that the existence of a CSR committee significantly increased ESG performance.⁵² The study differs from others in that it analyses two European countries in which the two-tier system is applied, which makes it particularly relevant for my study and it is to be expected that the results will be similar. However, the period analysed was more than 6 years ago compared to this study: changes in sustainability reporting and legislation on women on company boards may have had an impact, which could change these findings.

Horbach and Jacob (2018) find for German companies that there is a positive and significant relationship between mixed-gender boards and eco-innovation. This finding is explained by women having inspirational and transformative leadership styles; for innovation collaboration instead of competition is necessary.⁵³ Glass et al. (2016) analysed this relationship with Fortune 500 companies over a longer period of 10 years. They find that a high proportion of women does not directly lead to greater sustainability; however, if the female board members are also members of other companies' boards, they significantly enhance environmental practices.⁵⁴

A literature review conducted by Velte (2023) finds that the positive influence of board gender diversity on CSR performance can be seen both in countries that have voluntary board quota regimes and in countries that have fixed board quotas.⁵⁵ Of the three analysed variables board gender diversity, sustainability board expertise and sustainability-related executive compensation, gender diversity was found to be the most relevant in past research.⁵⁶ For the energy sector, as a CSR-sensitive industry, Shahbaz et al. (2020) analyse factors influencing ESG performance. They find that women on the board have a positive impact on overall ESG performance, as well as environmental and governance aspects. Surprisingly, however, there is no significant influence on the social dimension. They highlight the existence of CSR committees

as another factor that has a positive influence.⁵⁷

Birindelli et al. (2018) analyse the relationship between women on the board and ESG performance for banks in Europe and the United States. They find an inverted U-shaped relationship: up to the critical mass of three women or until 30 per cent of the members are female, ESG performance increases. After that, it decreases again or becomes insignificant with each additional woman on the board.⁵⁸ Nuber and Velte (2021) confirm this curvilinear relationship in their study but add that this could also be because the sample contains many companies with few women and few companies with many women on the board. Furthermore, they find evidence that there have to be at least two women on the board to be able to significantly reduce greenhouse gas emissions.⁵⁹

However, some studies come to different conclusions regarding the impact of women on ESG performance. Rao and Tilt (2021) find that while women bring objectivity and independence to the board, this does not necessarily translate to positive CSR outcomes, as there are barriers that hinder the positive relationship. These barriers include the low number of female board members and the limited support they receive from male board members.⁶⁰ This result does not show that women have a negative or no influence, but merely that the positive potential cannot be realised.

A study, with a primary focus on the influence of gender on stock market liquidity, shows that the representation of women on boards in France negatively influences participation in sustainable development projects.⁶¹

Although there are studies that find a mixed or negative impact, the majority of studies conclude that a higher proportion of women on the management board increases the company's ESG performance. For this reason, I formulate the following hypothesis:

Hypothesis 1 (H1): There is a positive association between the percentage of female management board members and ESG performance.

With regard to the influence that women on the management board have on sustainability, it is not only the relative proportion that is relevant, but it is also assumed that a certain absolute number of women must be represented on the management board to have a significant influence on ESG performance. This number is referred to as the critical mass. Critical mass theory can be traced back to Kanter (1977). She was the first to describe how individual women, in a male-dominated group, are merely tokens; a symbolic representation of their social category. Only when a certain critical mass is reached, women can have an impact in the group.⁶² Konrad et al. (2008) conducted interviews

⁵⁰ Cf. Zhang (2012), p. 695.

⁵¹ Cf. Harjoto et al. (2015), p. 641–642 for these sentences.

⁵² Cf. Velte (2016b), p. 107 for these sentences.

⁵³ Cf. Horbach and Jacob (2018), p. 931 for these sentences.

⁵⁴ Cf. Glass et al. (2016), p. 506–507 for these sentences.

⁵⁵ Cf. Velte (2022), p. 14

⁵⁶ Cf. Velte (2022), p. 17.

⁵⁷ Cf. Shahbaz et al. (2020), p. 11 for these sentences.

⁵⁸ Cf. Birindelli et al. (2018), p. 11–12 for these sentences.

⁵⁹ Cf. Nuber and Velte (2021), p. 176 for these sentences.

⁶⁰ Cf. Rao and Tilt (2021), p. 76–77 for these sentences.

⁶¹ Cf. Loukil et al. (2019), p. 698.

⁶² Cf. Kanter (1977), p. 966–968 for these sentences.

with women on the board of directors and used them to develop the critical mass hypothesis specifically for women on boards. This states that women can only make a difference once a critical mass of women has been reached on the board. One woman merely acts as a token, for example, to fulfil demands for female board participation. Although two women have more influence than one, they still have problems being heard, partly because they stand out among the male members. The critical mass is reached when there are three or more women on the board: they are seen as normal, support each other and work together towards common goals.⁶³

Yarram and Adapa (2021) also find a positive and significant influence of the percentage of female directors on the CSR score. The study, which looked at Australian companies in the ASX 300, confirms tokenism: this relationship does not hold for companies with only one female director. Only when there are two female directors the relationship becomes significant. Female board members show communal instead of agentic behaviour and support CSR concerns.⁶⁴

The finding of Birindelli et al. (2018), that ESG performance only increases until a threshold of three women on the board is reached and further women have no effect, does not support critical mass theory, in the narrow sense.⁶⁵ Rather, it proves the dual critical mass theory of Schwartz-Ziv (2017). The dual critical mass theory states that boards are most effective when genders are balanced and the board has at least three members from each gender.⁶⁶ These findings do not contradict the critical mass theory but emphasize that critical mass goes both ways; only increasing the participation of women does not necessarily increase ESG performance. This is also referred to as the too-much-of-a-good-thing effect, where after reaching an inflection point an input, previously having a desirable effect, leads to an unwanted outcome.⁶⁷

Bear et al. (2010) also find support for the critical mass theory, but only to the extent that CSR performance increases with an increasing number of women on the board. Although the statement is made that women are better able to assert themselves in a group, this is not statistically proven.⁶⁸ The fact that a critical mass is required to have a sizable effect can be observed not only in ESG performance but also in sustainability disclosure. Alkhwaja et al. (2023) find that a critical mass of three women is necessary to have a highly positive impact on ESG reporting. Although less than three women also have a significantly positive influence, it is far smaller.⁶⁹

The theory of critical mass is widely recognised and has also been confirmed by several studies, both in relation to ESG performance and beyond. Although the exact size of the critical mass certainly depends on the size of the board, qualitative and empirical literature most often cites three women

on the board as the threshold at which women leave minority status and have a significant impact. I therefore use this threshold and formulate the following hypothesis:

Hypothesis 2 (H2): There is a positive association between a critical mass of three or more women on the management board and ESG performance.

In comparison to the number of publications that exist on the relationship between board gender diversity and CSR, the influence of the CEO received less attention. This might be due to the relatively low number of female CEOs.⁷⁰ Busenbark et al. (2016) present three perspectives from which the CEO can be viewed: the position, the person and the environment.⁷¹ The first two aspects will be considered here. The CEO holds the position of the company's top decision-maker and can significantly influence strategic decisions. Crossland and Hambrick (2010) find that for Germany the CEO accounts for 11 per cent of variance in return on assets, for the United Kingdom even 19.5 per cent.⁷² To understand the actions of the CEO, the person and his or her individual characteristics must be considered. In this regard, Busenbark et al. (2016) again draw parallels to the upper echelons theory. CEOs identify strongly with the company and represent the company to stakeholders. The personality and characteristics of the CEO also have a major influence on firm outcomes.⁷³ While gender is not explicitly mentioned, it can be assumed that there are differences between men and women that are large enough to have an impact on organisational outcomes.

Aabo and Giorici (2023) examine the influence of female CEOs on ESG scores. They find a significant and positive association for Bloomberg ESG scores but not for ESG scores provided by Refinitiv, even when the companies in the sample are the same. The results indicate that it makes a difference which provider is used for the ESG scores.⁷⁴ For the UK setting Al-Shaer et al. (2024) find that a critical mass of female directors must be met for them to have a positive impact on ESG performance. However, female CEOs have a positive impact without the need to reach this critical mass of female directors. Young female CEOs with a short tenure have an especially positive influence.⁷⁵ Contrary to what one might assume, gender-diverse boards are particularly effective in emphasising environmental protection when the CEO is male.⁷⁶ The finding that gender diversity on boards with male CEOs is particularly important is supported by Liu (2018): environmentally damaging behaviour is less frequent on diverse boards. On the other hand, a female CEO only has a significant effect, when there are few women on the board.⁷⁷

⁶³ Cf. Konrad et al. (2008), p. 160 for these sentences.

⁶⁴ Cf. Yarram and Adapa (2021), p. 8 for these sentences.

⁶⁵ Cf. Birindelli et al. (2018), p. 12.

⁶⁶ Cf. Schwartz-Ziv (2017), p. 753.

⁶⁷ Cf. Pierce and Aguinis (2013), p. 331.

⁶⁸ Cf. Bear et al. (2010), p. 217.

⁶⁹ Cf. Alkhwaja et al. (2023), p. 12 for these sentences.

⁷⁰ Cf. Velte (2020), p. 1310.

⁷¹ Cf. Busenbark et al. (2016), p. 237.

⁷² Cf. Crossland and Hambrick (2010), p. 812.

⁷³ Cf. Busenbark et al. (2016), p. 248–251 for these sentences.

⁷⁴ Cf. Aabo and Giorici (2023), p. 5–6 for these sentences.

⁷⁵ Cf. Al-Shaer et al. (2024), p. 24 for these sentences.

⁷⁶ Cf. Glass et al. (2016), p. 506–507.

⁷⁷ Cf. Liu (2018), p. 137–138 for these sentences.

However, some studies indicate that no significant association between CEO gender and ESG performance exists. For the Indonesian banking sector, Sumarta et al. (2021) find no significant relationship. This finding might be explained by the limited number of female CEOs in the sample companies.⁷⁸ Glass et al. (2016) find a positive influence of female CEOs on environmental strengths and a negative influence on environmental weaknesses, but this relationship is not significant after controlling for other variables. In their sample, the proportion of female CEOs is only 2 per cent.⁷⁹ Although the low number of female CEOs is not explicitly mentioned by the authors as the reason for the lack of significance, it should be considered as a possible reason.

The findings on the influence of female CEOs on ESG performance are mixed. Whether a positive association is found depends, for example, on the ESG data used. Several studies, however, have found a positive association. Other studies conclude that female CEOs have no significant influence on ESG performance; in these studies, the proportion of female CEOs is consistently very low, which may also be the cause of this result. If one considers the upper echelons theory and sees the CEO as the individual who makes the company's key decisions, it can be assumed that women, with their character traits, have a significant influence on the sustainability of the company. I therefore formulate the following hypothesis:

Hypothesis 3 (H3): There is a positive association between the CEO being female and ESG performance.

Since in the two-tier system the management board is appointed by the supervisory board, there might exist spillover effects from women on supervisory boards to female members on management boards. If this were the case, FÜPoG I would also have an indirect effect on the number of women on the management board. This relationship is examined by Bozhinov et al. (2021) who find that indeed there exists a positive and significant relationship between women on the supervisory board and women on the management board. This is especially true if they are appointed by the shareholders (in contrast to being appointed by the employees) and they are serving on the nominating committee.⁸⁰ There is evidence that in companies that have more women on the supervisory board, there are also more women on the management board. However, it is questionable whether there exists a causal relationship, or other influences, such as increased pressure for the participation of women on boards, explain this effect.⁸¹

Dienes and Velte (2016) examine the impact of the percentage of women on the supervisory board on CSR disclosure and find a positive and significant relationship. Companies with a higher percentage of women on the supervisory board report more intensely on CSR topics and disclose

more information.⁸² In the context of mandatory sustainability reporting in Germany, Gerwing et al. (2022) identify a positive and significant association between the proportion of women on the supervisory board and the reporting quality. Surprisingly, this relationship does not hold to the proportion of women on the management board. One possible explanation for this is that less than 25 per cent of the companies analysed have a woman on their management board, and the overall proportion of women in these positions is low.⁸³

A study by HR consultancy Egon Zehnder (2021) in collaboration with the University of Göttingen describes a positive and significant association between the percentage of women on supervisory boards and corporate sustainability. The study analysed 534 companies, but the exact methodology and therefore the significance of the results remain unclear.⁸⁴

Due to the lack of previous (rigorous) studies dealing with the impact of gender diversity on the German supervisory board on ESG performance, it should be analysed to what extent the influence of women on the supervisory board differs from that of women on the management board and women in the one-tier system. From this, it can then be deduced whether a similar positive effect on ESG performance can be expected and whether the previous results can also be transferred to the supervisory board.

Zahra and Pearce II (1989) identify three roles that boards (of directors) assume: service, strategy and control. The service role involves establishing external contacts, providing the organisation with external legitimacy and advising and supporting management with expertise. The control role comprises the supervision, evaluation and monitoring of management.⁸⁵ The strategic role includes giving advice to management and proposing strategic changes.⁸⁶ In Germany, only the supervisory role of the supervisory board is legally recognised, while the other two roles are not. Based on these roles, Steger and Jahn (2019) conducted a survey of 116 German companies in 2006. In doing so, they attempt to close the research gap as to whether the supervisory board only has the statutory control role, for example, the appointment of board members, or additionally an active and advising role. Up to this point in time, the existing research is ambiguous on this matter.⁸⁷ They find that ethical issues play a major role in the work of the supervisory board, even more so than management tasks. Although the control role is considered the most important of the three, the respondents also attached great importance to the service and strategy roles. It can therefore not be said that the supervisory board is merely limited to monitoring the management board. The larger the board size, the more pronounced the service and strategy roles are.⁸⁸

⁷⁸ Cf. Sumarta et al. (2021), p. 1027 for these sentences.

⁷⁹ Cf. Glass et al. (2016), p. 503–504 for these sentences.

⁸⁰ Cf. Bozhinov et al. (2021), p. 1325–1326 for these sentences.

⁸¹ Cf. Kirsch and Wrohlich (2020), p. 49 for these sentences.

⁸² Cf. Dienes and Velte (2016), p. 14–15 for these sentences.

⁸³ Cf. Gerwing et al. (2022), p. 543 for these sentences.

⁸⁴ Cf. Egon Zehnder (Ed.) (2021), p. 3 for these sentences.

⁸⁵ Cf. Zahra and Pearce II (1989), p. 292–294 for these sentences.

⁸⁶ Cf. Zahra and Pearce II (1989), p. 298.

⁸⁷ Cf. Steger and Jahn (2019), p. 362–363 for these sentences.

⁸⁸ Cf. Steger and Jahn (2019), p. 369–370 for these sentences.

Based on these findings, it is possible to explain why the composition of the supervisory board might influence ESG performance. Sustainability as an ethical issue could be considered as important by the supervisory board and, therefore, be influenced by it. However, it is also to be expected that the results will differ from those of the management board and the existing findings in the one-tier system. In this respect, further evaluation is appropriate. I formulate the following hypothesis:

Hypothesis 4 (H4): There is a positive association between the percentage of female supervisory board members and ESG performance.

4. Research Design

4.1. Data and Sample

My initial sample consists of companies listed on the German stock indices DAX, MDAX and SDAX in the years 2020-2021. This is after the FÜPoG I but before FÜPoG II came into force. The three stock market indices consist of the 160 largest German companies on the Prime Standard of Deutsche Börse by free-float market capitalization.⁸⁹ To be included in the sample the company had to be part of the indices at least once during the 2-year period. Companies that were included for one day after a spin-off and afterwards no longer part of the indices are not included in the sample. This sample composition was selected because the majority (87.4 per cent in 2022) of the companies listed on the regulated market and that were subject to equal codetermination belonged to one of these indices.⁹⁰ These are also the companies to which the minimum participation requirements of FÜPoG II apply. In line with Velte (2016b) financial institutions are excluded from the sample.⁹¹ This is due to the fact, that financial institutions, like banks and insurance companies, often have a higher leverage⁹² and operate in a different regulatory environment. Except for the two variables *LogTA* and *ROA*, which were collected from the Wharton Research Data Service (WRDS), all variables were obtained from the Refinitiv (formerly known as Thomson Reuters) database. The number of women on the management board, the management board size and single missing data points that were not available on Refinitiv or WRDS were hand-collected from the companies' annual reports.

As the study is designed to obtain information specifically on the impact of women on the management board and supervisory board in the two-tier system, companies that do not fulfil this requirement are excluded from the sample. This applies in particular to the European legal form of the SE, where there is a free choice between a one- and two-tier.⁹³

As a result, 28 company-year observations are excluded. Furthermore, I exclude 27 observations with missing ESG data. The reason for missing ESG scores is often that the company no longer existed in the year or did not yet exist. The final sample consists of 305 firm-year observations from 157 distinct companies.

4.2. Variable Measurement

4.2.1. Independent Variables

To test H1, that is the effect of women on the management board on ESG performance, the independent variable *MBGend* is the percentage of women on the management board on 31 December of the year. It is calculated by dividing the number of women on the management board by board size. The independent variable for testing H4, *SBGend*, is calculated according to the same principle, with the difference that here the number of women on the supervisory board is divided by the size of the supervisory board. The independent variable for H3, *CEO*, is a dummy variable that is equal to one if the CEO of the company on 31 December was female and 0 if the CEO was male. H2 is tested using four dummy variables: *Women0*, *Women1*, *Women2* and *Women3*. *Women0* is equal to one if there are no female members in the management board and zero otherwise. *Women1*, *Women2* and *Women3* are equal to one if there are at least one, two or three female members on the management board, respectively, and zero otherwise. The regression is done four times with each dummy variable as the independent variable.

4.2.2. Dependent Variable

The dependent variable in all main models is *ESG*. ESG scores from the Refinitiv database are used as a proxy for ESG performance. This approach is in line with multiple earlier studies investigating the effect of women on boards on ESG performance.⁹⁴ Refinitiv's ESG score is based on more than 630 data points, assessing ESG performance on a scale from 0, being the lowest, to 100 being the highest attainable score. They are collected from companies' annual reports, company websites, stock exchange filings and other publicly available sources. Of these 630 data points the 186 most comparable for each industry are selected and make up the final ESG score. The score consists of a weighted average of three pillar scores 'Environment', 'Social' and 'Governance'. The weights differ for each industry. For the environmental subscore, the categories covered are *Emission*, *Innovation* and *Resource Use*. The social score consists of the categories *Community*, *Human Rights*, *Product responsibility* and *workforce*. The Governance score covers *CSR strategy*, *Management* and *Shareholders*. Refinitiv also provides an ESG controversy score, which is not part of the ESG score used in this study.⁹⁵

⁸⁹ Cf. STOXX (Ed.) (2024), p. 33.

⁹⁰ Cf. Frauen in die Aufsichtsräte e.V. (Ed.) (2022), p. 4.

⁹¹ Cf. Velte (2016b), p. 102.

⁹² Cf. Fama and French (1992), p. 429.

⁹³ Cf. Council Regulation (EC) No 2157/2001, Article 38.

⁹⁴ Cf. Aabo and Giorici (2023), p. 3; Bigelli et al. (2023), p. 5; Velte (2016b), p. 102.

⁹⁵ Cf. Refinitiv (Ed.) (2022), p. 4-10 for this paragraph.

4.2.3. Control Variables

Control variables are included in the regression models, to control for other factors that influence ESG performance. First, all independent variables, also act as control variables, because it can be assumed that these have an influence on the ESG score and are therefore included in all models. Only the critical mass dummy variables are not included when testing for the other hypotheses, since *MBGend* controls for the percentage of women on the management board more precisely. Conversely, in regression model 2, where the critical mass hypothesis is tested, the percentage of women on the management board is not included as a control variable, since they would be correlated. The percentage of women on the supervisory board is included in all models, the dummy variable for female CEOs is only included in model 3, where the CEO hypothesis is tested.

Previous research shows that firm and board characteristics influence ESG performance, which have to be controlled for and included in the statistical model. The selection of control variables included in the model is based on previous studies that examine the effects of women on ESG performance.

First, I control for the size of both the management board *MBSize* and the supervisory board *SBSize*. This is the number of members of each board at the end of the year. The size of the board has different influences on corporate governance. On the one hand, larger boards are less efficient at making decisions. This is because coordination and communication in larger groups of people is more difficult than in smaller groups.⁹⁶ This can lead to boards being unable to effectively implement strategic changes, which would be particularly necessary for a shift towards more sustainability.⁹⁷ On the other hand, Birindelli et al. (2018) argue that larger boards are more diverse, and offer a broader range of expertise and opinion.⁹⁸

The company's leverage or debt ratio is included in the model and calculated by dividing the company's debt by the total assets to form the variable *DebtRatio*. In previous studies, it is very common to include the debt ratio as an independent variable in the model. However, the direction of the effect of high leverage on ESG performance is not clear and some authors find a positive association⁹⁹, while others find a negative relationship¹⁰⁰. This can be explained by two different theories. On the one hand, companies with a high level of debt are very dependent on their creditors, who therefore represent an important stakeholder group. This can lead to the interests of other stakeholders, such as society, and thus the commitment to sustainability, being neglected. On the other hand, high debt is associated with increased agency costs, which can be reduced through sustainability and sus-

tainability disclosure practices.¹⁰¹ In both cases, it can be assumed that the debt ratio explains part of ESG performance and is therefore included in my model.

As a measure of firm size, the natural logarithm of total assets *LogTA* is used. The natural logarithm is taken to reduce the skewness of the variable. This is common practice in finance literature.¹⁰² Gallo and Christensen (2011) find that company size has a positive and significant influence on the sustainability behaviour of companies. This is because larger companies have more resources at their disposal to deal with sustainability issues and are more in the public eye.¹⁰³

The control variable *ROA* is used to measure financial performance by including the firm's Return on Assets in the model. This is calculated by dividing income before extraordinary items by total assets. Again, arguing with slack resources theory, Waddock and Graves (1997) find that firms with more slack resources, resulting from better financial performance, can spend more of these resources to improve their sustainability. Furthermore, there also seems to be evidence that the causality runs in the reverse direction; financial performance is improved by engaging in sustainability activities which is referred to by the authors as a 'virtuous circle'.¹⁰⁴

SBMeetings controls for the number of supervisory board meetings in a given year. Meeting frequency data for the management board is not available and thus cannot be controlled for. Evidence for the influence of board meeting frequency is mixed. One study suggests that there is a positive and significant influence of the number of board meetings on sustainable development and stakeholder interest.¹⁰⁵ Frequency is positively associated with sustainability disclosure, implying that firms with more board meetings have more effective CSR strategies.¹⁰⁶ Vafeas (1999) finds the opposite is true; high board meeting frequency is associated with poor financial performance, suggesting that board meeting frequency increases in times of crisis.¹⁰⁷ Applied to sustainability, this would mean that a high number of supervisory board meetings could indicate that the company is in distress and thus also reduce its ESG performance, as fewer resources can be allocated to it. For the German setting, no significant influence of supervisory board meeting frequency on CSR reporting intensity was found.¹⁰⁸

SBIIndependence indicates the ratio of independent members to the total number of supervisory board members. According to the German Corporate Governance Code (2022), a supervisory board member is considered independent if he or she has no personal or business relationship with the company or the management board which could result in a conflict of interest. This is particularly the case if the supervisory

⁹⁶ Cf. Huther (1997), p. 263 for these sentences.

⁹⁷ Cf. Goodstein et al. (1994), p. 248.

⁹⁸ Cf. Birindelli et al. (2018), p. 5.

⁹⁹ Cf. Al-Shaer et al. (2024), p. 13.

¹⁰⁰ Cf. Velte (2016b), p. 107.

¹⁰¹ Cf. Khaled et al. (2021), p. 3 for these sentences.

¹⁰² Cf. Dang et al. (2018), p. 161.

¹⁰³ Cf. Gallo and Christensen (2011), p. 36 for these sentences.

¹⁰⁴ Cf. Waddock and Graves (1997), p. 314 for these sentences.

¹⁰⁵ Cf. Dube and Jaiswal (2015), p. 175.

¹⁰⁶ Cf. Jizi (2017), p. 651.

¹⁰⁷ Cf. Vafeas (1999), p. 140-141.

¹⁰⁸ Cf. Dienes and Velte (2016), p. 12.

board member or their close relative has been on the supervisory board for more than 12 years, has been a member of the management board in the last two years, is a close relative of a member of the management board or has had a significant business relationship with the company itself or with a company dependent on the company in the last year.¹⁰⁹ Research shows that board independence is positively associated with the support of sustainable development goals¹¹⁰ and ESG scores¹¹¹ and negatively with greenhouse gas emissions¹¹². This is because independent supervisory board members are in a better position to monitor the activities of the management board, are more focused on stakeholder interests and are less focused on achieving short-term objectives¹¹³, with sustainability being a long-term goal¹¹⁴.

CSRCommittee is a dummy variable that indicates the existence of a CSR Committee and is equal to one if the company has a CSR committee and zero otherwise. It is a sub-committee of either the management or the supervisory board, to increase corporate social performance and create transparency with regard to sustainability issues.¹¹⁵ Previous research shows that the implementation of a CSR committee has a positive and significant effect on environmental and social scores, but no significant effect on governance in the German setting.¹¹⁶ Based on this finding and in line with previous research, which finds a positive and significant effect on ESG scores¹¹⁷, this variable is included in the model.

SusCompIncen is a dummy variable equal to 1 if the members of the management board are compensated with regard to ESG performance and 0 otherwise. Firms that implement such compensation regimes are, on average, more likely to have better CSR performance. CSR-linked compensation is more common in firms with independent boards and better corporate governance. CSR is not seen as an agency cost but rather beneficial to financial performance and thereby to shareholders.¹¹⁸

To account for differences between years and industries I include year and industry fixed effects. *YEAR* is a dummy variable equal to 0 for the fiscal year 2022 (t-1: 2021) and equal to one for the fiscal year 2021 (t-1=2020). The industry fixed effects dummy variables are based on the first two digits of the Global Industry Classification Standard (GICS), accounting for 11 industry sectors (10 after excluding financial institutions). They are equal to 1 if the company belongs to that industry sector and 0 otherwise. The energy industry is chosen as the reference point, and is therefore not included in the regression models.

The models does not include a control variable that con-

trols for whether the company falls under the provisions of FÜPoG II. There are several reasons for this: firstly, it is not expected that whether a company is subject to the law has a direct impact on its ESG performance as measured by the ESG score. Rather the effect of the FÜPoG II on ESG performance is mediated by the number of women in the management board, which is included in the regression. Secondly, it is conceivable that the introduction of such a variable could lead to multicollinearity, as the number of women on boards could in turn depend on whether the law is applied. Furthermore, the majority of the data is from the period before the introduction of FÜPoG II. For these reasons, I decided against the inclusion of such a variable.

4.3. Regression Model

In line with Velte (2016b),¹¹⁹ I use three linear OLS regression models to test the hypotheses. To examine the impact of female management board members and female supervisory board members on ESG performance and test H1 and H4 I use the following model:

$$\begin{aligned} \text{ESG} = & \alpha + \beta_1 \text{MBGend} + \beta_2 \text{SBGend} + \beta_3 \text{MBSize} \\ & + \beta_4 \text{SBSize} + \beta_5 \text{DebtRatio} + \beta_6 \text{LogTA} \\ & + \beta_7 \text{ROA} + \beta_8 \text{SBMeetings} + \beta_9 \text{SBIndependence} \quad (1) \\ & + \beta_{10} \text{CSRCommittee} + \beta_{11} \text{SusCompIncen} \\ & + \text{Year FE} + \text{Industry FE} + \varepsilon \end{aligned}$$

α denotes the intercept, β s are the regression coefficients and ε is the error term.

H1 and H4 are tested using the same model, as it can be expected from previous research that the percentage of women on the management board as well as on the supervisory board influences the ESG score, therefore they both explain part of the variance. The dummy variable *CEO* for the CEO being female is not included, since the CEO is part of the management board. Therefore, companies with a female CEO have more women on the management board, which makes *MBGend* and *CEO* not statistically independent and increases the risk of multicollinearity, would they be included in the same model. By adding industry and time-fixed effects, I control for the differences between the two observed years and between industries.

H2, whether a critical mass of women on the management board is necessary to affect ESG performance, is tested using dummy variables that are equal to 1 if a certain absolute number of women on the board is reached and 0 otherwise as the independent variable. Regression model 2 is estimated four times, each time including only one of the independent variables. This determines the impact on ESG performance if a certain threshold of women as members of the management board is met. *MBGend* is no longer used to control for the proportion of women on the management

¹⁰⁹Cf. DCGK C recommendation C.6-7 for these sentences.

¹¹⁰Cf. Tagliatela et al. (2023), p. 2499.

¹¹¹Cf. Bigelli et al. (2023), p. 11.

¹¹²Cf. Kim et al. (2023), p. 13.

¹¹³Cf. Zaid et al. (2020), p. 1356.

¹¹⁴Cf. Aragón-Correa and Sharma (2003), p. 84.

¹¹⁵Cf. Eberhardt-Toth (2017), p. 1926–1927.

¹¹⁶Cf. Baraibar-Diez and Odriozola (2019), p. 16.

¹¹⁷Cf. Birindelli et al. (2018), p. 14; Velte (2016b), p. 107.

¹¹⁸Cf. Hong et al. (2016), p. 205 for these sentences.

¹¹⁹Cf. Velte (2016b), p. 104.

board in this model as it would not be statistically independent from the critical mass dummy variables. Again, the *CEO* variable is not included in that model, since it would be correlated with the dummy variables. All control variables that were included in regression model 1 are also included in this model.

$$\begin{aligned} \text{ESG} = & \alpha + \beta_1 \text{Women0} + \beta_2 \text{Women1} + \beta_3 \text{Women2} \\ & + \beta_4 \text{Women3} + \beta_5 \text{SBGend} + \beta_6 \text{MBSize} \\ & + \beta_7 \text{SBSsize} + \beta_8 \text{DebtRatio} + \beta_9 \text{LogTA} + \beta_{10} \text{ROA} \\ & + \beta_{11} \text{SBMeetings} + \beta_{12} \text{SBIndependence} \\ & + \beta_{13} \text{CSRCommittee} + \beta_{14} \text{SusCompIncen} \\ & + \text{Year FE} + \text{Industry FE} + \varepsilon \end{aligned} \quad (2)$$

The regression model that is used for testing H3 is the only model, that additionally contains the *CEO* variable, which is the independent variable in this model. *SBGend* is included as a control variable in the model. To avoid multicollinearity issues, *MBGend* is not included in this model; since CEOs are part of the management board, naturally the percentage of female members on the management board will be higher if the CEO is female. This is in line with other studies investigating the effect of female CEOs on ESG performance.¹²⁰ Control variables are the same as in regression models 1 and 2, as it can be expected that the factors influencing the ESG score are the same.

$$\begin{aligned} \text{ESG} = & \alpha + \beta_1 \text{CEO} + \beta_2 \text{SBGend} + \beta_3 \text{MBSize} \\ & + \beta_4 \text{SBSsize} + \beta_5 \text{DebtRatio} + \beta_6 \text{LogTA} + \beta_7 \text{ROA} \\ & + \beta_8 \text{SBMeetings} + \beta_9 \text{SBIndependence} \\ & + \beta_{10} \text{CSRCommittee} + \beta_{11} \text{SusCompIncen} \\ & + \text{Year FE} + \text{Industry FE} + \varepsilon \end{aligned} \quad (3)$$

Linear OLS regression has multiple assumptions which are tested here. They are: normal distribution of the error term, no multicollinearity among the independent variables, homoscedasticity of residues, and linearity.¹²¹ Plotting the residuals it can be seen that there is no perfect linear relationship, but where the most data points are, the line is almost vertical at zero. Checking for normality of the distribution of residuals it can be observed that there is a normal distribution which is slightly negatively skewed. To detect multicollinearity of the independent variables, the variance inflation factor (VIF) is calculated. Except for the industry dummy variables, where some multicollinearity can be expected, the VIFs for regression equation 1 range from 1.1961 to 2.9264 with the mean being 1.5234. For regression equations 2 and 3 the VIFs are very similar, as mostly the same dependent variables are included in the model. VIF values larger than 10 are considered to be problematic and imply multicollinearity.¹²² The

VIF values in all models are well below the threshold, which indicates no problematic multicollinearity exists. Lastly, to test for homoscedasticity of residue a Breusch-Pagan-Test is used. The null hypothesis of the test is heteroscedasticity.¹²³ The p-value is below the 1 per cent significance level for all models which means I reject the null hypothesis, which in turn means, heteroscedasticity is likely to be present in the data.¹²⁴ Despite this, a linear regression model is used, because the other assumptions are met.

All independent variables are lagged by one year, to account for the delay in the causal effect. The ESG score is compared to the board composition of the previous year. This is because board members who are new to the board need time to settle in and contribute their ideas and suggestions. It also takes time for management decisions, that impact ESG performance, to have an effect. This approach is in line with multiple other studies that analyse the influence of gender diversity on sustainability.¹²⁵

5. Empirical Analysis

5.1. Descriptive Statistics

The descriptive statistics (mean, standard deviation, median, maximum and minimum) are presented in Table 1. The average ESG Score is 0.617, with 0 being the lowest and 1 being the highest attainable score. This is considerably higher than the mean of 0.251, that Velte (2016a) reports for companies belonging to the three DAX indices for the period 2010-2014.¹²⁶ Companies in the sample are rated at a wide range of scores, with the lowest score being 0.115 and the highest awarded score at 0.932. The mean of the environmental and social pillars of the ESG score *EnvSocMean* (without the governance pillar) is 0.608. That is only slightly lower than the ESG score itself and it has approximately the same minimum and maximum as the ESG score and also a similar standard deviation. The average management board has 4 members, with the smallest management board in the sample consisting of 2 and the largest of 9 members. On average only 13.9795 per cent of the management board members are female. The highest proportion of females in the sample is 50 per cent, which means there is no board on which there are more women than men. In 59 per cent of the boards, there is not a single female member. The position of the CEO is heavily male-dominated, with only 3.9344 per cent of the companies having a female CEO. There are only 12 firm-year observations with a female CEO. The proportion of women as CEOs is therefore 3.5 times lower than the proportion of women on the management board. In 44.26 per cent of the company-year observations, the remuneration of the management board members was linked to the ESG performance

¹²³Cf. Breusch and Pagan (1979), p. 1288.

¹²⁴See Appendix B 'Regression Assumptions' for detailed results for VIFs and Breusch-Pagan-Test

¹²⁵Cf. Menicucci and Paolucci (2022), p. 9-10; Velte (2016b), p. 103.

¹²⁶Cf. Velte (2016a), p. 21.

¹²⁰Cf. Aabo and Giorici (2023), p. 3; Al-Shaer et al. (2024), p. 8.

¹²¹Cf. Poole and O'Farrell (1971), p. 148.

¹²²Cf. Neter et al. (1983), p. 392.

Table 1: Descriptive Statistics

Variable	Mean	SD	Min	Median	Max
ESG	0.6169	0.1760	0.1156	0.6331	0.9323
EnvSocMean	0.6080	0.2020	0.0118	0.6229	0.9373
MBGend	0.1031	0.1398	0.0000	0.0000	0.5000
SBGend	0.2901	0.1345	0.0000	0.3333	0.6667
MBWomen	0.4590	0.6170	0.0000	0.0000	3.0000
SBWomen	3.3869	2.3483	0.0000	4.0000	9.0000
CEO	0.0393	0.1947	0.0000	0.0000	1.0000
MBSize	4.0000	1.5623	2.0000	4.0000	9.0000
SBSize	10.7475	5.5160	2.0000	12.0000	24.0000
DebtRatio	0.2620	0.1667	0.0001	0.2454	1.0329
LogTA	22.1734	1.7166	18.6200	21.9600	26.9900
LogMarketCap	22.0054	1.4921	18.8515	21.8318	25.7511
ROA	0.0294	0.0640	-0.2433	0.0362	0.2487
SBMeetings	8.0623	4.9477	3.0000	7.0000	52.0000
SBIndependence	0.5282	0.3561	0.0000	0.5000	1.0000
CSRCommittee	0.7836	0.4125	0.0000	1.0000	1.0000
SusCompIncen	0.4426	0.4975	0.0000	0.0000	1.0000
Women0	0.5967	0.4914	2.0000	1.0000	1.0000
Women1	0.4033	0.4914	0.0000	0.0000	1.0000
Women2	0.0459	0.2096	0.0000	0.0000	1.0000
Women3	0.0098	0.0989	0.0000	0.0000	1.0000

Notes: For definitions of the variables, please see Appendix A.

of the company. This means around half of the companies offer monetary incentives to increase ESG activity. The majority of companies have a CSR committee, namely 78.36 per cent. The supervisory board has an average of 10.7475 members, much larger than the average management board. The smallest supervisory board only has two and the largest has 24 members. The average proportion of women on the supervisory board is 29 per cent. This is far higher than the proportion of women on the management board. This is to be expected, as most of the companies in the sample fall under the provisions of FöPoG I, which stipulates a 30 per cent proportion of women by law. However, the average suggests that most companies do not appoint women to the supervisory board beyond the statutory minimum. While in 32 company-year observations, not a single woman is represented on the supervisory board, the 50 per cent mark is only exceeded in 4 firm-year observations.

On average, the supervisory board meets 8.06 times a year, with a very wide range of 3 to 52 meetings. The average proportion of independent supervisory board members is 52.82 per cent, which means that just over half of the supervisory board members are independent. There are also

supervisory boards that are completely independent and supervisory boards that do not have a single independent member. The average firm in the sample has total assets of € 4.26 billion. The log of total assets is taken to create the variable *LogTA*. The mean return on assets in the sample is 2.94 per cent and the mean debt ratio is 0.262.

Table 2 presents the Pearson correlation matrix for all variables but the year and industry dummy variables. The results do not suggest any serious multicollinearity in the regression. This is also supported by the low VIFs, as discussed in section 4.3. As expected, the dependent variable *ESG* and the independent variable *MBGend* are positively and significantly correlated, however, the correlation is not very strong (0.2122). The independent variable for the proportion of women on the supervisory board is slightly stronger correlated with *ESG* (0.3247) and the correlation is highly significant at the 1 per cent level. *ESG* is slightly negatively (-0.0284) correlated with the *CEO* dummy, however this correlation is not significant. All independent variables and control variables except *CEO*, *DebtRatio*, *ROA*, *SBMeetings* are significantly correlated with *ESG*. Except for the insignificant *CEO* variable, all correlations are positive.

Table 2: Pearson Correlation Matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) ESG	1										
(2) EnvSockMean	0.9345***	1									
(3) MBGend	0.2122**	0.1967***	1								
(4) SBGend	0.3247***	0.3003***	0.0327	1							
(5) MBWomen	0.3260	0.3154***	0.8704***	0.0721	1						
(6) SBWomen	0.4436***	0.4847***	0.1596***	0.7019***	0.2970***	1					
(7) CEO	-0.0284	-0.0294	0.3117***	-0.1007*	0.1777***	-0.0262	1				
(8) MBSIZE	0.3656***	0.3836***	0.2139***	0.1527***	0.5460***	0.4205***	-0.0649	1			
(9) SBSIZE	0.4177***	0.4901***	0.1523***	0.3679***	0.3057***	0.8753***	0.0062	0.4546***	1		
(10) DebtRatio	0.0773	0.0535	0.1384**	0.0788	0.1890***	0.1121*	-0.0353	0.0507	0.1063*	1	
(11) LogTA	0.5676***	0.5678***	0.2417***	0.2891***	0.4435***	0.6132***	-0.0323	0.5219***	0.6652***	0.2568***	1
(12) LogMarketCap	0.5846***	0.5417***	0.1761***	0.2013***	0.3825***	0.4289***	-0.0613	0.4924***	0.4310***	0.0501	0.7753***
(13) ROA	0.0265	0.0255	0.0267	-0.1534***	0.0311	-0.1606***	0.0089	0.0553	-0.1498***	-0.2668***	-0.148***
(14) SBMeetings	0.0751	0.0494	0.0809	-0.0761	0.0606	-0.0312	0.0043	0.0281	-0.0273	0.1547***	0.1344**
(15) SBIndependence	0.3415***	0.2175***	-0.0131	0.0819	0.0150	0.0015	0.0102	0.0239	-0.0590	0.1685***	0.1908***
(16) CSRCommittee	0.4614***	0.4488***	0.1486***	0.1343**	0.1848***	0.2090***	0.1063*	0.1378**	0.2347***	0.1197**	0.3390***
(17) SusComplnec	0.2540***	0.1852***	0.2374***	0.1469**	0.2682***	0.2077***	0.0573	0.1270**	0.1775***	0.0334	0.2708***
(18) Women0	-0.3091***	-0.3038***	-0.8987***	-0.0825	-0.9064***	-0.2834***	-0.2462***	-0.4628***	-0.2938***	-0.1406**	-0.3951***
(19) Women1	0.3091***	0.3038***	0.8987***	0.0825	0.9064***	0.2834***	0.2462***	0.4628***	0.2938***	0.1406**	0.3951***
(20) Women2	0.1678***	0.1501***	0.3638***	-0.0213	0.6250***	0.1175**	-0.0444	0.4018***	0.1352**	0.1589***	0.2656***
(21) Women3	0.1425**	0.1404**	0.1942***	0.0855	0.4111***	0.1961***	-0.0202	0.2556***	0.1614***	0.1439**	0.2408***
(12) LogMarketCap	1										
(13) ROA	0.1315**	1									
(14) SBMeetings	0.0553	-0.2525***	1								
(15) SBIndependence	0.2469***	0.0112	0.1061*	1							
(16) CSRCommittee	0.2606***	-0.0433	0.1404**	0.2495***	1						
(17) SusComplnec	0.2863***	-0.0104	0.0329	0.0915	0.0996*	1					
(18) Women0	-0.3337***	-0.0191	-0.0708	-0.0006	-0.2210***	-0.2363***	1				
(19) Women1	0.3337***	0.0191	0.0708	0.0006	0.2210***	0.2363***	-1***	1			
(20) Women2	0.2479***	0.0510	-0.0091	0.0104	0.0011	0.1831***	-0.2668***	0.2668***	1		
(21) Women3	0.2029***	-0.0087	0.0458	0.0686	0.0524	0.1118*	-0.1212**	0.1212**	0.4544***	1	

Notes: *, **, *** represent significance levels of 0.10, 0.05, and 0.01, respectively. For definitions of the variables, please see Appendix A.

5.2. Regression Results and Discussion

Table 3 shows the results for the regression model using *MBGend* and *SBGend* as independent variables, investigating the relationship between females on the supervisory and management board on ESG performance (H1 and H4). As a goodness of fit measure, the adjusted R^2 is employed. The adjusted R^2 of the model is 0.546, indicating a good fit of the regression model. Between *MBGend* and *ESG* I find a positive and significant association, with a regression coefficient of 0.1136 at the 5 per cent significance level. This suggests that for a 1 per cent increase in women on the management board, the *ESG* score increases by 0.1136 points. Compared to a company with no women on the management board, a company that has 50 per cent women on the management board, the *ESG* score would be 5.68 points higher. For the interpretation, it should be noted that although the variable *ESG* is coded on a scale of 0 to 1, the *ESG* score itself is given on a scale of 0 to 100. The results are in support of H1 and suggest that a higher proportion of women on the management board positively affects ESG performance. The direction of the effect is in line with earlier studies, that show that female members and gender diversity on the management board have a positive and significant influence on ESG performance.¹²⁷ Velte (2016b) also finds a positive and significant association for the German setting, however, the effect size he finds is more than twice as large, with a coefficient of 0.254, which suggests that for every 1 per cent increase in women on the management board, the *ESG* score increases by 0.254 points. Furthermore, the association Velte finds is stronger, with significance at the 1 per cent level.¹²⁸ This difference could be explained by the different sample used, which in addition to German companies also consisted of Austrian companies and the fact that an earlier period was examined.

With the average management board only consisting of 4 members and the average proportion of females on the management board being only 10.31 per cent, the results suggest that although there is a relatively low proportion of female management board members in the sample companies, they do have a significant impact on ESG performance. This indicates that a critical mass of female management board members is not needed to have a significant and positive influence on ESG performance and that even few women can make a positive impact. These results are in contrast to Rao and Tilt (2021), who found that although women have a positive influence on sustainability in principle, the low number of women on the board and the associated barriers prevent this positive influence from unfolding.¹²⁹ My results reflect what the majority of previous research has found. They show that women, even in small numbers as in the sample, can have a significant positive impact on sustainability. This matter will be analysed in more detail using regression model 2,

which uses dummy variables to investigate how the impact of women on ESG performance changes, if the number of women on the management board increases.

Table 3: Female Supervisory and Management Board Members Regression

Independent Variable	ESG	
	Coefficient	p-value
Intercept	-0.4096	0.0039***
MBGend	0.1136	0.0336**
SBGend	0.1802	0.0023***
MBSize	0.0053	0.3373
SBSize	0.0023	0.2304
DebtRatio	-0.0461	0.3226
LogTA	0.0396	0.0000***
ROA	0.1205	0.3282
SBMeetings	0.0009	0.5438
SBIndependence	0.1236	0.0000***
CSRCommittee	0.1068	0.0000***
SusCompIncen	0.0288	0.0572*
Year Effects	included	
Industry Effects	included	
N	305	
Adj. R^2	0.546	

Notes: *, **, *** represent significance levels of 0.10, 0.05, and 0.01, respectively. For definitions of the variables, please see Appendix A.

The regression results for regression model 2 are displayed in Table 4. It tests H2, namely that a critical mass of three or more women on the management board is needed to have a significant effect on ESG performance. For the hypothesis to be accepted, the increase in the coefficient between *Women2* and *Women3* should be significantly higher than the increase from *Women1* to *Women2*. Furthermore, according to tokenism theory the impact of one woman, represented by the coefficient of *Women1*, on the management board should be insignificant. The regression is run 4 times, each time with a different independent dummy variable, which represents a threshold of an absolute number of women on the management board being met. The coefficient -0.0308 for *Women0* is negative and significant at the 10 per cent level. This suggests that having no women on the management board has a negative and significant effect on ESG performance. As expected, having at least one woman on the management board has the opposite effect of having no women on the management board; the sign of the coefficient for *Women1* is switched in comparison to *Women0*, the p-value is the same. This implies that including one woman on the management board has a significant and positive impact on ESG performance. For

¹²⁷Cf. Harjoto et al. (2015), p. 641–642; Bear et al. (2010), p. 217; Shahbaz et al. (2020), p. 11.

¹²⁸Cf. Velte (2016b), p. 106 for these sentences.

¹²⁹Cf. Rao and Tilt (2021), p. 76–77.

Table 4: Critical Mass Regression

Independent Variable	ESG							
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Intercept	-0.3679	0.0120**	-0.3987	0.0053***	-0.4128	0.0059***	-0.4320	0.0026***
Woman0	-0.0308	0.0723*						
Woman1			0.0308	0.0723*				
Woman2					0.0437	0.2556		
Woman3							0.0235	0.7549
SBGend	0.1803	0.0024***	0.1803	0.0024***	0.1808	0.0025***	0.1721	0.0038***
MBSize	0.0032	0.5854	0.0032	0.5854	0.0044	0.4511	0.0063	0.2621
SBSize	0.0022	0.2474	0.0022	0.2474	0.0023	0.2275	0.0022	0.2609
DebtRatio	-0.0436	0.3496	-0.0436	0.3496	-0.0428	0.3636	-0.0365	0.4364
LogTA	0.0395	0.0000***	0.0395	0.0000***	0.0397	0.0000***	0.0404	0.0000***
ROA	0.1299	0.2922	0.1299	0.2922	0.1373	0.2670	0.1429	0.2485
SBMeetings	0.0010	0.4954	0.0010	0.4954	0.0011	0.4790	0.0010	0.5101
SBIndependence	0.1228	0.0000***	0.1228	0.0000***	0.1203	0.0000***	0.1187	0.0000***
CSRCommittee	0.1058	0.0000***	0.1058	0.0000***	0.1109	0.0000***	0.1098	0.0000***
SusCompIncen	0.0302	0.0461**	0.0302	0.0461**	0.0326	0.0312**	0.0343	0.0227**
Year Effects	included		included		included		included	
Industry Effects	included		included		included		included	
N	305		305		305		305	
Adj. R ²	0.5439		0.5439		0.5408		0.5389	

Notes: *, **, *** represent significance levels of 0.10, 0.05, and 0.01, respectively. For definitions of the variables, please see Appendix A.

the third dummy variable *Women2*, the coefficient is slightly higher, at 0.0437, however it is insignificant at a p-value of 0.2556. For *Women3* the coefficient is the lowest of the three at 0.0235. Again it is not significant at a p-value of 0.7549. These results are opposite to what would have been necessary to accept H2. The statistical analysis suggests that the first woman on the management board has a positive and significant impact on ESG performance, and that while additional women further improve sustainability, a single woman also makes a major contribution. A critical mass of three women is not needed to have a positive and significant effect on ESG performance.

The conclusion that can be drawn from the data is that not having a woman on the management board has a negative impact on ESG performance. This means that the *ESG* score is on average 3.2 points lower than in companies that have at least one woman on the management board.

The results of regression model 2 also emphasise the results of model 1 and confirm H1, namely that women on the board have a positive influence on ESG performance. The coefficients for *Women2* and *Women3* are not significant, which is why they should be interpreted with caution. However, the

interpretation of the coefficients would allow the conclusion to be drawn about an inverted U-shaped relationship; the coefficient of *Women2* is greater than the coefficient of *Women3*, which means that the positive influence on ESG performance decreases again when there are more than 2 women on the management board. These results would be in line with previous research¹³⁰, but are based on non-significant results so further research is necessary to confirm these findings.

Part of the critical mass theory also states that a single woman is merely a token woman and has no positive impact on sustainability. The fact that there are so many companies with only one woman in the sample and that both *MBGend* in regression model 1 and regression model 2 with the independent variable *Women1* have a positive and significant effect on the sustainability of the company, allow the conclusion that even a single woman on the board can have a positive effect on the sustainability of the company. This is particularly interesting in the context of FÜPoG II, as the law only obliges companies to appoint one woman to the management board.

¹³⁰Cf. Birindelli et al. (2018), p. 11–12; Nuber and Velte (2021), p. 1976.

However, another explanation for the regression results, namely that including a second and third woman on the management board does not lead to a larger increase in ESG performance than the inclusion of the first women, should also be considered: there are only three firm-year observations in the sample that have three women on the board, two of which belong to the same firm. The influence of at least three women on the board on ESG performance is therefore only determined based on these three observations. Also, there are only 14 observations in which 2 or more women are present on the management board. If the dummy variables *Women2* and *Women3* only take the value one in such a small number of observations, the regression coefficient for the independent variable will most likely not be significant, even if there were a significant effect in the population. Although the model does not confirm the critical mass theory, conversely it cannot be said that the critical mass theory is rejected. There are just too few observations in the sample in which the hypothesised critical mass of women is reached to be able to come to a meaningful conclusion about whether the critical mass effect does or does not exist. To conduct a meaningful study with this sample, the number of companies with at least 3 women on the board would have to be significantly higher. Although this regression model again confirms that more women generally have a positive influence on the sustainability of companies, there are too few women on the management boards to be able to make statements about the validity of the critical mass theory.

Regarding the question of whether critical mass holds true for the German setting, my results neither confirm that the critical mass theory is correct, nor can they clearly refute it, due to my sample containing too few companies in which the critical mass is reached. However, one important contribution can be made regarding token women. By describing token women, who have little influence on strategic processes and are included on the board mainly for outward appearances, previous literature often gives the impression that having one woman is as good as having none.¹³¹ However, this would be a serious error, as my results show that a single woman can also have a positive influence and that the inclusion of a woman is the first step, albeit a small one, towards a board with equal representation.

Table 5 shows the results for regression model 3 which tests for H3, whether there is a positive impact on female CEOs on ESG performance, as measured by the *ESG* score. The coefficient of the *CEO* variable is negative at -0.0189. However, the relationship is not statistically significant, as indicated by a p-value of 0.6056. The lack of significance means H3 can be rejected, implying there is no significant relationship between female CEOs and ESG performance. This result is not necessarily because female CEOs have no impact on ESG performance. In fact, there are multiple explanations for such a result. The first, being the most obvious, is that female CEOs indeed have no effect on ESG performance. Another possible explanation is, that the hypothesis is true for

the population, but there is just not enough evidence in the sample to support the hypothesis, also referred to as a Type II error.¹³² The most probable cause of this error for my sample is, that the sample size is too small, the characteristic hypothesized to have an effect, that is the CEO being female, occurs too rarely. Since only 12 of the total 305 firm-year observations have a female CEO, even if there were a significant effect, it may not be measurable due to the small number of female CEOs. To rule out with certainty that female CEOs have no influence on ESG performance, a sample would be required in which the proportion of female CEOs is higher or the total number of observations is larger. It can be concluded that for my sample of German DAX, MDAX and SDAX firms in the two-year period observed, no significant effect of the female CEOs on ESG performance can be found.

Table 5: Female CEO Regression

Independent Variable	ESG	
	Coefficient	p-value
Intercept	-0.4316	0.0025***
CEO	-0.0189	0.6056
SBGend	0.1695	0.0045***
MBSize	0.0065	0.2468
SBSize	0.0023	0.2448
DebtRatio	-0.0363	0.4374
LogTA	0.0403	0.0000***
ROA	0.1435	0.2462
SBMeetings	0.0010	0.5065
SBIndependence	0.1188	0.0000***
CSRCommittee	0.1108	0.0000***
SusCompIncen	0.0350	0.0205**
Year Effects	included	
Industry Effects	included	
N	305	
Adj. R ²	0.5391	

Notes: *, **, *** represent significance levels of 0.10, 0.05, and 0.01, respectively. For definitions of the variables, please see Appendix A.

Based on the upper echelon theory and the associated assumption that the characteristics of top management have a significant influence on corporate decisions and thus also on sustainability, and the findings that women on management and supervisory boards have a positive influence on sustainability, it is surprising that this relationship should not exist for female CEOs. Especially as the CEO holds the position of top decision maker in the company.¹³³

¹³²Cf. Visentin et al. (2020), p. 918.

¹³³Cf. Busenbark et al. (2016), p. 237.

¹³¹Cf. Kanter (1977), p. 966–968; Konrad et al. (2008), p. 160.

However, these results are in line with part of the previous research, which also found a non-significant association of female CEOs on ESG performance when using ESG scores specifically provided by Refinitiv. The insignificant nature of my finding could also be due to the data provider that provides the ESG scores as a significant positive relationship was found for Bloomberg ESG scores, but not for ESG scores provided by Refinitiv.¹³⁴ My results also match the findings of Sumarta et al. (2021) who also find a non significant relationship and use the same explanation, namely that there are few female CEOs in their sample.¹³⁵

In support of H4, I find a positive and significant association between *SBGend* and *ESG* with a regression coefficient of 0.1802 which is significant at the 1 per cent level. This result suggests that women on the supervisory board have a positive and significant impact on ESG performance. In comparison to the effect women have on the management board, the effect of women on the management board is larger, as the coefficient of *SBGend* is more than 1.5 times higher than the coefficient of *MBGend*. This is rather surprising as members of the management board are more directly involved in managing the company's affairs and its strategic positioning. This would suggest that the positive effect women have, according to the upper echelons theory, would be larger for the management board. But the regression results suggest that the opposite is the case.

To date, there have been no other studies in the German setting that use a regression model to analyse the impact of women on the supervisory board on ESG performance. On the one hand, my results are therefore not comparable with previous study results for the German Setting. On the other hand, however, they provide new and important insights into how women on supervisory boards affect ESG performance. The fact that the gender of supervisory board members, and thus their characteristics according to the upper echelons theory, have an influence on sustainability as a management outcome of the company also ties in with previous findings on the changing role of the supervisory board.¹³⁶ It shows that the supervisory board not only has a controlling function but that it also plays a role that leads to strategic changes within the company and therefore has an influence on ethical and social issues such as sustainability. My results are in line with previous research on the impact of women on boards and as expected. They are consistent with findings that women on boards in general have a positive impact on ESG performance. They also complement the research in the German setting that the proportion of women on the supervisory board has a positive influence not only on sustainability reporting¹³⁷ but also on ESG performance itself.

As the regression models largely contains the same control variables, these are considered together in the following section and their impact on ESG performance is described

and analysed.

The coefficients for both board size variables, *MBSize* for the management board size and *SBSize* for the supervisory board size are insignificant with a p-value of 0.3373 and 0.2304 respectively, for regression model 1. As previous research suggests, board size has no clear effect on ESG performance, as larger boards are, on the one hand, more inefficient at communicating and coordination,¹³⁸ and on the other hand, offer a broader range of diversity and opinions.¹³⁹ This is also reflected in the insignificant association I find in the data, as the direction of the effect is unambiguous and the results are expected.

The association between *DebtRatio* and *ESG* is found to be insignificant at a p-value of 0.3226 for regression model 1 and a slightly higher p-value for models 2 and 3. As expected from previous studies the coefficient is negative at -0.0461. Findings from previous studies are inconclusive with some studies showing a positive effect of high leverage and others showing a negative effect on ESG performance. As discussed earlier, this can be due to reducing agency costs associated with high debt through sustainable actions or the neglect of the community stakeholders by focusing mainly on creditors, when debt is high, leading to lower ESG performance.¹⁴⁰ As the effect of high leverage can lead to ESG performance going in both directions for different companies, or even within the same company, this can lead to the association being insignificant.

In all regression models, there is a positive and significant association between the natural logarithm of *LogTA* and *ESG*. This suggests that firm size has a positive influence on ESG performance. The regression coefficient in model 1 is 0.0396 and does not deviate in the other models. It is significant at the 1 per cent level. Because of the log transformation of the independent variable, the interpretation of the regression coefficient is not straightforward. The results suggest that if the total assets of a firm increase by 1 per cent the *ESG* score increases by 0.0394 points.¹⁴¹ The finding that ESG performance increases with increasing firm size, as proxied by total assets, is consistent with previous studies.¹⁴² This was expected as larger firms have more resources available to invest in sustainable development and practices.

The association between *ROA* and *ESG* is positive with a coefficient of 0.1205, however it is insignificant. Previous research would have suggested a positive but significant relationship.¹⁴³

In all models the control variable *SBMeetings* is insignificant, in model 1 the p-value is 0.5438, in models 3 it is also above the 50 per cent significance level. The regression coefficient is close to zero (0.0009) for model 1. This result is not unexpected, as the effects and reasons for a large number of supervisory board meetings are not unambiguous. On

¹³⁸Cf. Goodstein et al. (1994), p. 248.

¹³⁹Cf. Birindelli et al. (2018), p. 5.

¹⁴⁰Cf. Khaled et al. (2021), p. 3.

¹⁴¹ $\beta_6 = 0.0396$, $\beta_6 * \ln(1.01) = 0.000394$

¹⁴²Cf. Chang et al. (2024), p. 14; Liu (2018), p. 126; Velte (2016b), p. 106.

¹⁴³Cf. Gallo and Christensen (2011), p. 336.

¹³⁴Cf. Aabo and Giorici (2023), p. 5–6 for these sentences.

¹³⁵Cf. Sumarta et al. (2021), p. 1027.

¹³⁶Cf. Steger and Jahn (2019), p. 369–370.

¹³⁷Cf. Dienes and Velte (2016), p. 14–15.

the one hand, a high number may indicate difficulties in the company or inefficiencies in management; on the other hand, many meetings may also indicate better and more intensive communication and diligence. The finding of an insignificant association between the frequency of board meetings and the ESG score is in line with previous research, which also reports a positive coefficient that is not significant.¹⁴⁴

SBIIndependence is positively and significantly associated with ESG at the 1 per cent significance level. The regression coefficient is 0.1236 for model 1, suggesting that if the percentage of independent members on the supervisory board rose by 1 per cent, the ESG score would increase by 0.1253 points. This finding was expected, as independent board members are more likely to monitor management activities more objectively, and reflects what was found in previous studies.¹⁴⁵

As expected, the association between *CSRCommittee* and ESG is positive and significant at the 1 per cent significance level. The coefficient is 0.1068 implying that a company that has a CSR Committee has an ESG score that is on average 10.68 points higher, than a company that has not implemented a CSR Committee. The results for regression models 2 and 3 also show a significant and positive association with similar coefficients. However, it should also be taken into account that other effects may play a role here. For example, companies that place a particularly high value on sustainability are more likely to implement a CSR committee, meaning that the committee is not the cause of the increased ESG performance, but rather that there is a reverse causality.

I find that there exists a positive association between the existence of sustainability compensation incentives and ESG performance. The association is significant at the 10 per cent level for regression model 1 and for regression model 3 even at the 5 per cent level. The regression coefficient for *SusCompIncen* in model 1 is 0.0288, implying that for companies that have linked management board remuneration to the achievement of sustainability targets, the ESG score is on average 2.88 points higher, compared to companies that have no sustainability compensation incentives. Although the association is positive and significant the effect is relatively small. This is in line with previous research, which found that integrating CSR criteria into executive compensation leads to improved ESG performance, by increasing management long-term orientation, increasing environmental and social initiatives, increasing environmental innovation and reducing greenhouse gas emissions.¹⁴⁶

Year and industry dummies were included in all three models. The year dummy was significant at the 10 per cent level for model 1 and 2 and insignificant for model 3.

One limitation of this study is that the operationalisation of ESG performance was based on Refinitiv's ESG scores. It should be noted that there are ESG scores available from other providers, like Sustainalytics or MSCI. These scores

significantly differ from each other, with the main driver of this divergence being how the ESG score is measured. Furthermore, there is a rater effect, meaning that if a company receives a good score in one category, it also receives good scores in other categories.¹⁴⁷ These findings should be taken into account when interpreting the regression results. Chatterji et al. (2016) find that ESG scores from different raters have a low correlation and that they vastly differ from one another, although raters are measuring the same theoretical construct, that is ESG performance.¹⁴⁸ Depending on which ESG score is used, different associations between board characteristics and ESG performance can be found. Aabo and Giorici (2023) find no association between the CEO gender and ESG performance when using ESG scores provided by Refinitiv, but when using ESG scores by the data provider Bloomberg a positive and significant association can be found. This is mainly due to differences in the environmental and social scores.¹⁴⁹ It would be interesting to repeat the study with ESG data from other data providers, like Bloomberg and see if the outcomes match or are different.

Furthermore, the data covers only a period of two years. During these two years, changes, especially those that occurred due to legislation, are difficult to observe. This would require more than two years of data. While the impact of women on the management board is important to understand the potential effect of FÜPoG II, the law only went into effect in August of 2022. This period is only covered by the lagged ESG variable and only by half a year. The effects of FÜPoG II are not contained in the data, especially for the number of women on the management board since the independent variables are from the years 2020 and 2021. To address this limitation, another study would have to be carried out covering a longer period both before and after the introduction of the law. This is not possible with the data currently available, as ESG data is available for 2022 at the latest. In my research so far, I have assumed that the influence of women on ESG performance is linear; more women lead to more sustainability. This is also reflected in my choice of a linear regression model. However, it is also conceivable that the underlying mechanism is not the proportion of women, which increases sustainability, but gender diversity. This would mean that from a certain proportion of women, for example, 50 per cent, the ESG performance decreases again, as male perspectives are missing. In this case, there would be an inverted U-shaped relationship in which the ESG performance first increases with an increasing proportion of women, reaches its maximum at an unknown point and then decreases again, as found in previous studies.¹⁵⁰ Regression model 2 provides an indication of this relationship, but it is based on non-significant coefficients. To investigate this relationship further, it would be necessary to have a sample in which all distributions of men and women occur, including

¹⁴⁷Cf. Berg et al. (2022), p. 1341 for these sentences.

¹⁴⁸Cf. Chatterji et al. (2016), p. 1607–1608.

¹⁴⁹Cf. Aabo and Giorici (2023), p. 5 for these sentences.

¹⁵⁰Cf. Birindelli et al. (2018), p. 11–12; Nuber and Velte (2021), p. 1976.

¹⁴⁴Cf. Al-Shaer et al. (2024), p. 13; Birindelli et al. (2018), p. 11.

¹⁴⁵Cf. Al-Shaer et al. (2024), p. 13.

¹⁴⁶Cf. Flammer et al. (2019), p. 1097.

companies where women are overrepresented. In my sample of DAX, MDAX and SDAX companies, the highest proportion of women on the management board is 50 per cent. There are no management boards in the sample, in which men are in the minority. In 4 firm-year observations, there are more women on the supervisory board than men. Again, not enough to analyse what happens when women outnumber men. However, it is probable that the positive trend in the number of women on supervisory boards and management boards will continue, making it possible to analyse this matter in the future.

A further limitation results from the sample size in combination with the fact that the proportion of women on the management board and in the CEO position is relatively low. Although the results regarding the critical mass and CEO hypotheses are insignificant, a significant association cannot be ruled out. This is because a small sample size reduces the statistical power of the regression and therefore a true effect that potentially exists cannot be detected.¹⁵¹ This may change if more women are represented on the management board, also because of FüPoG II. However, the law only requires that a woman be represented on the management board if there are at least three members. It is therefore questionable whether the number of companies with four or more women would increase. This would be necessary to adequately test the critical mass hypothesis. Possibly, an increase in the number of women on the board would also increase the number of female CEOs, which would in turn improve the situation for testing the CEO hypothesis (H3). Another possibility would be to extend the analysis to a larger number of companies and thus increase the number of women in the sample.

6. Robustness Test and Further Analysis

In the following section, I will carry out several robustness checks. I will test the same hypotheses but change some model specifications and check whether I obtain the same results. Regression model 1 is addressed in particular, as this was the focus of the work and delivered significant results. I will test whether these significant results persist when the model specifications are changed. In addition, interaction effects are used to analyse the conditions under which the influence of women on the supervisory board and management board on ESG performance changes or is particularly pronounced.

One issue with the Refinitiv ESG score and my regression model is that the calculation for the governance subscore also included how diverse the management is, how the compensation of the management is structured, and which committees exist in the company. These factors are also included in my regression models as the dependent variables *MBGend* and *SBGend*, which indicate the proportion of women on the management and supervisory board but also as control variables such as *CSRCommittee* and *SusCompIncen*. As these factors are included in the calculation of the ESG score itself, it

is quite clear that there must exist a significant relationship between these control variables and the dependent variable *ESG*. The following section analyses whether and to what extent the percentage of women has an impact on sustainability if these factors are not included in the variable that approximates ESG performance. Furthermore, excluding the governance score from the dependent variable addresses endogeneity and reverse causality issues.¹⁵²

Therefore, in the first robustness test, the dependent variable *EnvSocMean* only includes the environment and social subscores. This is done by calculating the mean of both scores as provided by Refinitiv. The governance score is no longer part of the dependent variable.

The following regression model is estimated:

$$\begin{aligned} \text{EnvSocMean} = & \alpha + \beta_1 \text{MBGend} + \beta_2 \text{SBGend} \\ & + \beta_3 \text{MBSize} + \beta_4 \text{SBSize} \\ & + \beta_5 \text{SBMeetings} + \beta_6 \text{SBIndependence} \\ & + \beta_7 \text{CSRCommittee} \\ & + \beta_8 \text{SusCompIncen} + \beta_9 \text{LogTA} \\ & + \beta_{10} \text{DebtRatio} + \beta_{11} \text{ROA} \\ & + \text{Year FE} + \text{Industry FE} + \varepsilon \end{aligned} \quad (4)$$

The adjusted R^2 is 0.4974, which is 0.0487 lower than the original model. This was to be expected, as some of the independent variables are no longer included in the regression and therefore the model has lower explanatory power. Both the coefficients of *MBGend* and *SBGend* are still positive and there is a significant association at the 10 per cent level for *MBGend* and the 5 per cent level for *SBGend*. The effect size remains largely the same with coefficients for *MBGend* at 0.1069 and *SBGend* at 0.1587. This suggests that the results of the main model are not due to the board's diversity being included in the calculation of the ESG score, but rather that women on both the supervisory board and the management board do indeed have a positive impact on sustainability.

A large difference can be seen in the control variable *SusCompIncen*, which controls for the sustainability-oriented remuneration of the management board. This was positive in the main model and significant at the 10 per cent level. The p-value in this model is 0.5217, indicating that the relationship between sustainability compensation incentives and the mean of sustainability and social scores is no longer significant. This sharp drop in significance due to the exclusion of compensation in the dependent variable is rather surprising as previous research did not find a significant relationship between sustainability related compensation and the governance score for the German setting.¹⁵³ Based on this research it would have been expected that the relationship between *SusCompIncen* and *EnvSocMean* would have become even more significant after the exclusion of the governance score, but the opposite was the case.

¹⁵¹Cf. Button et al. (2013), p. 365.

¹⁵²Cf. Bigelli et al. (2023), p. 10.

¹⁵³Cf. Baraibar-Diez et al. (2019), p. 1469.

Table 6: Excluded Governance Score Regression

EnvSocMean		
Independent Variable	Coefficient	p-value
Intercept	-0.3410	0.0455**
MBGend	0.1069	0.0971*
SBGend	0.1587	0.0257**
MBSize	0.0066	0.3241
SBSize	0.0070	0.0027***
DebtRatio	-0.0660	0.2410
LogTA	0.0388	0.0000***
ROA	0.1502	0.3130
SBMeetings	0.0002	0.8992
SBIndependence	0.0802	0.0026***
CSRCommittee	0.1268	0.0000***
SusCompIncen	0.0116	0.5217
Year Effects	included	
Industry Effects	included	
N	305	
Adj. R ²	0.4974	

Notes: *, **, *** represent significance levels of 0.10, 0.05, and 0.01, respectively. For definitions of the variables, please see Appendix A.

The second robustness test deals with the operationalisation of the independent variable. In line with the majority of existing research, the percentage of women on the entire board was used in the main model. However, another approach would be to use the number of women on the board as an independent variable. The dependent variable is, as in the main model, *ESG*. The following regression model is used:

$$\begin{aligned}
 \text{ESG} = & \alpha + \beta_1 \text{MBWomen} + \beta_2 \text{SBWomen} + \beta_3 \text{MBSize} \\
 & + \beta_4 \text{SBSize} + \beta_5 \text{SBMeetings} \\
 & + \beta_6 \text{SBIndependence} + \beta_7 \text{CSRCommittee} \\
 & + \beta_8 \text{SusCompIncen} + \beta_9 \text{LogTA} \\
 & + \beta_{10} \text{DebtRatio} + \beta_{11} \text{ROA} \\
 & + \text{Year FE} + \text{Industry FE} + \varepsilon
 \end{aligned} \quad (5)$$

The regression model shows that the results are robust to changes to the operationalisation of the independent variables. There is a positive and significant association between the number of women, both on the management and the supervisory board, and the *ESG* score. For both variables, it is significant at the 10 per cent level. Here, the coefficient for *MBWomen* is 0.0246, which is about twice as large as the coefficient for *SBWomen*, which is 0.0119. In the main model, it was the exact opposite; the influence of women on the supervisory board was twice as high as that of women on the

management board. This could be due to the fact that the supervisory board is generally larger: the same increase in absolute terms, leads, in relative terms, to a smaller increase in the supervisory board compared to the management board.

Table 7: Absolute Number of Women Regression

ESG		
Independent Variable	Coefficient	p-value
Intercept	-0.3485	0.0172**
MBWomen	0.0246	0.0918*
SBWomen	0.0119	0.0654*
MBSize	0.0013	0.8284
SBSize	-0.0007	0.8218
DebtRatio	-0.0471	0.3206
LogTA	0.0393	0.0000***
ROA	0.1011	0.4157
SBMeetings	0.0006	0.6797
SBIndependence	0.1257	0.0000***
CSRCommittee	0.1097	0.0000***
SusCompIncen	0.0301	0.0502*
Year Effects	included	
Industry Effects	included	
N	305	
Adj. R ²	0.5351	

Notes: *, **, *** represent significance levels of 0.10, 0.05, and 0.01, respectively. For definitions of the variables, please see Appendix A.

The third robustness test is intended to check whether changes in the control variables lead to the same regression result. The results are displayed in Table 8. For this purpose, variables are omitted from the models or other proxies are used.

First, I omit all control variables from the model and only use the variables *MBGend* and *SBGend* as explanatory variables. This is to estimate a baseline of the model, before including any other factors that influence *ESG* performance. The following regression equation is used:

$$\text{ESG} = \alpha + \beta_1 \text{MBGend} + \beta_2 \text{SBGend} + \varepsilon \quad (6)$$

Both coefficients remain positive and significant. The p-value for both coefficients is much smaller compared to the main model, and the coefficients are larger. This indicates that there is still a significant association between women on boards and *ESG* performance without controlling for any other factors and that adding control variables decreases the significance of the two independent variables. However, the fit of the model without control variables is very poor, with an

Table 8: Alternate Control Variables Regression

Independent Variable	ESG					
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Intercept	0.46995	0.0000***	-0.4225	0.0016**	-0.7651	0.0000***
MBGend	0.2541	0.0002***	0.1257	0.0175**	0.0890	0.1441
SBGend	0.4163	0.0000***	0.1889	0.0012***		
MBSize			0.0054	0.3207	0.0037	0.5579
SBSize			0.0050	0.0035***		
DebtRatio			-0.0003	0.9953	-0.0164	0.7591
LogTA					0.0604	0.0000***
LogMarketCap			0.0397	0.0000***		
ROA			-0.0635	0.6095	0.1415	0.3043
SBMeetings			0.0008	0.5844		
SBIndependence			0.1101	0.0000***		
CSRCommittee			0.1111	0.0000***		
SusCompIncen			0.0233	0.1235	0.0336	0.0539*
Year Effects	excluded		included		included	
Industry Effects	excluded		included		included	
N	305		305		305	
Adj. R ²	0.1405		0.5554		0.3955	

Notes: *, **, *** represent significance levels of 0.10, 0.05, and 0.01, respectively. For definitions of the variables, please see Appendix A.

R² of 0.1405, which is significantly improved, when adding more control variables.

In model 7 I use a different firm size proxy: the natural logarithm of the market capitalization *LogMarketCap*. This is also a commonly used measure of firm size in financial modelling¹⁵⁴ and is therefore well suited to replace the total assets in the model. The following model is used:

$$\begin{aligned}
 \text{ESG} = & \alpha + \beta_1 \text{MBGend} + \beta_2 \text{SBGend} + \beta_3 \text{MBSize} \\
 & + \beta_4 \text{SBSize} + \beta_5 \text{SBMeetings} \\
 & + \beta_6 \text{SBIndependence} + \beta_7 \text{CSRCommittee} \\
 & + \beta_8 \text{SusCompIncen} + \beta_9 \text{LogMarketCap} \\
 & + \beta_{10} \text{DebtRatio} + \beta_{11} \text{ROA} \\
 & + \text{Year FE} + \text{Industry FE} + \varepsilon
 \end{aligned} \tag{7}$$

The results show that the coefficients and significance levels for the independent variables of interest *MBGend* and *SBGend* are very similar to the main model. The results are therefore robust to changes in the firm size proxy. There has been one noteworthy change in the results. The coefficient for the size of the supervisory board *SBSize* is now significant at the 1 per cent level. The coefficient (0.0050) implies

that each additional member of the supervisory board would increase the ESG score by an average of 0.50 points.

Thirdly, I estimate a model, that excludes all variables that are related to the supervisory board, to test the robustness of the regression results regarding the independent variable *MBGend*. These variables are *SBGend*, *SBSize*, *SBMeetings*, *SBIndependence* and *CSRCommittee*:

$$\begin{aligned}
 \text{ESG} = & \alpha + \beta_1 \text{MBGend} + \beta_2 \text{MBSize} \\
 & + \beta_3 \text{SusCompIncen} + \beta_4 \text{LogTA} \\
 & + \beta_5 \text{DebtRatio} + \beta_6 \text{ROA} \\
 & + \text{Year FE} + \text{Industry FE} + \varepsilon
 \end{aligned} \tag{8}$$

The results for model 8 show that the coefficient for *MBGend* is still positive but no longer significant. However, the p-value for *MBGend* is 0.1441, which is above the 10 per cent significance level. The R² of 0.3955 is decreased which indicates that important explanatory variables, compared to the main model which had an R² of 0.5461, have been removed. The robustness test therefore indicates that the supervisory board variables improve the fit of the model and should not be omitted. The inclusion of the supervisory board variables is also justified as the activities of the management board and supervisory board are closely related.

¹⁵⁴Cf. Dang et al. (2018), p. 1.

To further analyse the relationship between the independent variables and how they interact with each other to affect ESG performance, I add interaction effects between the independent variables and between the independent variables and the control variables. This leads to a better understanding of the conditions under which women on boards have a particularly positive influence and the conditions under which this influence may be reduced. Table 9 displays the coefficients and p-values for all 21 combinations of independent and control variables, including two interaction terms with the independent variables and the year dummy variable. Each regression contains the interaction term, and all independent and control variables, used in regression model 1. Two of the interaction terms have a statistically significant coefficient. The statistically significant terms are *SBGend*DebtRatio* and *SBGend*CSRCommittee*. The other 19 interaction terms are insignificant. There is no significant interaction term that contains *MBGend*. Also, the interaction of the two independent variables *MBGend*SBGend* is insignificant with a p-value of 0.5717. This suggests that the percentage of women on one board does not change the impact the women on the other board have. This could have been the case, for example if women on management boards and supervisory boards had supported each other.

In regression model 9 the interaction term *SBGend*DebtRatio* is negative (-0.5687) and significant at the 10 per cent level. The following regression equation was used to estimate model 9:

$$\begin{aligned}
 \text{ESG} = & \alpha + \beta_1 \text{SBGend} * \text{DebtRatio} + \beta_2 \text{MBGend} \\
 & + \beta_3 \text{SBGend} + \beta_4 \text{MBSIZE} + \beta_5 \text{SBSIZE} \\
 & + \beta_6 \text{DebtRatio} + \beta_7 \text{LogTA} + \beta_8 \text{ROA} \\
 & + \beta_9 \text{SBMeetings} + \beta_{10} \text{SBIndependence} \\
 & + \beta_{11} \text{CSRCommittee} + \beta_{12} \text{SusCompIncen} \\
 & + \text{Year FE} + \text{Industry FE} + \epsilon
 \end{aligned} \tag{9}$$

This suggests that the higher the debt ratio of the firm, the lower the impact of women on the supervisory board on ESG performance. This could be possibly explained by the fact that companies with higher levels of debt are in a worse economic situation and women therefore have fewer opportunities to realise their positive effects on sustainability. For example, there is a lack of financial resources to implement sustainability measures or changes. The interaction between these two variables with influence on sustainability has not yet been described in the literature and should be investigated in more detail in the future.

Regression model 10 shows for the interaction term *SBGend*CSRCommittee* a positive coefficient at 0.2195 which is significant at the 5 per cent level with a p-value of 0.0458.

Table 9: Interaction Terms

Interaction Terms	ESG	
	Coefficient	p-value
MBGend*SBGend	-0.2069	0.5460
MBGend*MBSIZE	0.0060	0.8711
MBGend*SBSIZE	-0.0115	0.2243
MBGend*DebtRatio	0.3243	0.2869
MBGend*LogTA	0.0140	0.6501
MBGend*ROA	-0.9770	0.2783
MBGend*SBMeetings	0.0075	0.4759
MBGend*SBIndependence	-0.0862	0.5528
MBGend*CSRCommittee	-0.0466	0.7184
MBGend*SusCompIncen	0.0260	0.8003
MBGend*Year	0.0084	0.9331
SBGend*MBSIZE	-0.0487	0.2262
SBGend*SBSIZE	-0.0175	0.1398
SBGend*DebtRatio	-0.5687	0.0744*
SBGend*LogTA	0.0274	0.4989
SBGend*ROA	0.7742	0.3870
SBGend*SBMeetings	0.0000	0.5793
SBGend*SBIndependence	0.1476	0.3293
SBGend*CSRCommittee	0.2195	0.0459**
SBGend*SusCompIncen	-0.0553	0.6351
SBGend*Year	0.0156	0.8807

Notes: *, **, *** represent significance levels of 0.10, 0.05, and 0.01, respectively. For definitions of the variables, please see Appendix A.

The regression equation is:

$$\begin{aligned}
 \text{ESG} = & \alpha + \beta_1 \text{SBGend} * \text{CSRCommittee} + \beta_2 \text{MBGend} \\
 & + \beta_3 \text{SBGend} + \beta_4 \text{MBSIZE} + \beta_5 \text{SBSIZE} \\
 & + \beta_6 \text{DebtRatio} + \beta_7 \text{LogTA} + \beta_8 \text{ROA} \\
 & + \beta_9 \text{SBMeetings} + \beta_{10} \text{SBIndependence} \\
 & + \beta_{11} \text{CSRCommittee} + \beta_{12} \text{SusCompIncen} \\
 & + \text{Year FE} + \text{Industry FE} + \epsilon
 \end{aligned} \tag{10}$$

These results imply that if a company has a CSR committee, women on the supervisory board have a higher impact on ESG performance. Since the CSR committee is a sub-committee of the supervisory board it is likely that when there is a higher percentage of women on the supervisory board, there are also more women on the CSR committee. The CSR committee would then act as an outlet for the female members to increase their positive influence and focus on policies and strategies positively affecting the company's ESG performance. Interestingly the coefficient for *SBGend* is no longer significant in model 10.

Table 10: Significant Interaction Terms Regressions

Independent Variable	ESG			
	Coefficient	p-value	Coefficient	p-value
Intercept	-0.4295	0.0024***	-0.3877	0.0061***
SBGend*DebtRatio	-0.5687	0.0744*		
SBGend*CSRCommittee			0.2195	0.0458**
SBGend	0.3146	0.0011***	0.0450	0.6138
DebtRatio	0.1115	0.2632	-0.0373	0.4231
CSRCommittee	0.1006	0.0000***	0.0469	0.1823
MBGend	0.1154	0.0303**	0.1060	0.0467**
MBSize	0.0042	0.4477	0.0051	0.3579
SBSize	0.0028	0.1542	0.0022	0.2499
LogTA	0.0397	0.0000***	0.0394	0.0000***
ROA	0.1276	0.2991	0.1295	0.2912
SBMeetings	0.0009	0.5369	0.0011	0.4564
SBIndependence	0.1334	0.0000***	0.1216	0.0000***
SusCompIncen	0.0298	0.0488*	0.0292	0.0527*
Year Effects		included		included
Industry Effects		included		included
N		305		305
Adj. R ²		0.5495		0.5508

Notes: *, **, *** represent significance levels of 0.10, 0.05, and 0.01, respectively. For definitions of the variables, please see Appendix A.

7. Conclusion

In my analysis, I examined what impact women on boards have on ESG performance in the German two-tier system. For a sample of 157 DAX, MDAX and SDAX companies over a period of 2 years, I find that the presence of women on management boards as well as supervisory boards increases ESG performance. This relationship applies to both the percentage of women and the absolute number of women on the boards.

Further, the theory that women only have an impact on ESG performance once a critical mass of women has been reached has also been tested. However, no evidence to support that theory could be found in the data. In contrast to previous literature, I showed that even the inclusion of a single women, often referred to as a ‘token women’ has a significant and positive impact on ESG performance.

The analysis of the data failed to confirm that having a female CEO has a significant impact on ESG performance. However, it is not appropriate to conclude that female CEOs do not have a positive impact on sustainability or that the critical mass theory does not hold for German companies. Rather, there are too few CEOs and too few management board members in the sample companies who have reached the hypothetical critical mass of women to be able to make

a statement about how this affects the sustainability of the company in question.

Two important findings emerge from this study: firstly, that women in management positions can greatly improve the sustainability of a company. Secondly, the proportion of women in management positions, both as CEOs and as members of the management board, is significantly under-represented. The situation is more favourable on supervisory boards, but here often only the legal minimum is met. With the introduction of FÜPoG I and FÜPoG II, the first steps were taken to increase the proportion of women on supervisory boards to 30 per cent and on management boards with more than three members to at least one woman. However, these laws only apply to a small number of German companies that fulfil certain size criteria. In recent years, there has also been an increase in the number of women on these committees. In this respect, the FÜPoG I and FÜPoG II not only strengthen the position of women at the management level but are also expected to contribute to sustainability. If the critical mass theory is correct, which can be assumed according to the current state of research and which cannot be refuted in the context of my study, further regulation is required. With the FÜPoG II only requiring companies to include one woman on the management board, the proportion of women, especially on large

management boards, will not be significantly increased. The FÜPoG II thus goes just so far as to ensure that a woman is on the board when a certain board size is reached; although I found that even one woman has a positive effect on sustainability, critical mass theory that was shown to be true in previous research should not be disregarded. It would therefore be desirable to increase the minimum quota for women on management boards further, to include more women in top management positions.

Two main limitations could not be addressed as part of the robustness tests. Firstly, the small sample size and thus the small number of female CEOs and companies with several women on the board. An observation period of more than two years would also have been desirable, particularly following the entry into force of FÜPoG II. Secondly, the high subjectivity of the operationalisation of ESG performance through the ESG scores. Here, ESG scores from other data providers, such as Bloomberg, could have been used as a dependent variable to check whether the results are robust to this change in variable measurement. However, as I did not have this data available, this robustness test was not performed.

Companies must face up to the challenges posed by climate change and develop sustainable strategies to remain competitive and environmentally and socially responsible in the long term. On the one hand, research clearly shows that women have a positive influence on many financial and non-financial performance metrics and that companies should therefore be motivated to appoint more women to management positions. On the other hand, it also shows that legislative action is necessary to increase the proportion of women, as in many cases companies do not appoint women to the management and supervisory boards beyond the legal minimum. Further research is needed into the conditions that must be in place for the positive effect of women to unfold. This includes, for example, the association between increasing debt levels and the decreasing positive effect of women on the supervisory board on sustainability, which was found in this study. Gaining further insights could also lead to an increase in the proportion of women in companies, as the findings should motivate those responsible to utilise the positive influence that women have to the benefit of the company's shareholder and stakeholders.

References

- Aabo, T., & Giorici, I. C. (2023). Do Female CEOs Matter for ESG Scores? *Global Finance Journal*, 56.
- Adams, R. B., & Funk, P. (2012). Beyond the Glass Ceiling: Does Gender Matter? *Management Science*, 58(2), 219–235.
- Ahmad, S., & Omar, R. (2016). Basic Corporate Governance Models: A Systematic Review. *International Journal of Law and Management*, 58(1), 73–107.
- AktG. (2024). Aktiengesetz of 06 September 1965, as amended on 01 January 2024.
- Alkhwaja, A., Hu, F., Johl, S., & Nadarajah, S. (2023). Board Gender Diversity, Quotas, and ESG Disclosure: Global Evidence. *International Review of Financial Analysis*, 90, Article 102823.
- Al-Shaer, H., Zahman, M., & Albitar, K. (2024). CEO Gender, Critical Mass of Board Gender Diversity and ESG Performance: UK Evidence. *Journal of Accounting Literature*, ahead-of-print.
- Aragón-Correa, J., & Sharma, S. (2003). A Contingent Resource-Based View of Proactive Corporate Environmental Strategy. *The Academy of Management Review*, 28(1), 71–88.
- Arayssi, M., Dah, M., & Jizi, M. (2016). Women on Boards, Sustainability Reporting and Firm Performance. *Sustainability Accounting, Management and Policy Journal*, 7(3), 376–401.
- Bannò, M., Filippi, E., & Trento, S. (2023). Women in Top Echelon Positions and their Effects on Sustainability: A Review, Synthesis and Future Research Agenda. *Journal of Management and Governance*, 27(1), 181–251.
- Baraibar-Diez, E., & Odriozola, M. D. (2019). CSR Committees and their ESG Performance in UK, France, Germany, and Spain. *Sustainability*, 11(18), Article 5077.
- Baraibar-Diez, E., Odriozola, M. D., & Fernández Sánchez, J. L. (2019). Sustainable Compensation Policies and its Effect on Environmental, Social, and Governance Scores. *Corporate Social Responsibility & Environmental Management*, 26(6), 1457–1472.
- BDO AG Wirtschaftsprüfungsgesellschaft & Kirchhoff Consult AG (Ed.) (2023). Nachhaltigkeit in heißen Zeiten. Retrieved October 10, 2024, from https://www.kirchhoff.de/fileadmin/static/pdfs/2023_DAX160_Studie_gesamt.pdf
- Bear, S., Rahman, N., & Post, C. (2010). The Impact of Board Diversity and Gender Composition on Corporate Social Responsibility and Firm Reputation. *Journal of Business Ethics*, 97(2), 207–221.
- Berg, F., Kölbl, J. F., & Rigobon, R. (2022). Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance*, 26(6), 1315–1344.
- Bigelli, M., Mengoli, S., & Sandri, S. (2023). ESG Score, Board Structure and The Impact of the Non-Financial Reporting Directive on European Firms. *Journal of Economics and Business*, 127, Article 106133.
- Birindelli, G., Dell'Atti, S., Iannuzzi, A. P., & Savioli, M. (2018). Composition and Activity of The Board of Directors: Impact On ESG Performance In The Banking System. *Sustainability*, 10(12), Article 4699.
- Bozhinov, V., Joecks, J., & Scharfenkamp, K. (2021). Gender Spillovers from Supervisory Boards to Management. *Managerial and Decision Economics*, 42(5), 1317–1331.
- Breusch, T., & Pagan, A. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5), 1287–1294.
- Busenbark, J. R., Krause, R., Boivie, S., & Graffin, S. D. (2016). Toward a Configurational Perspective on the CEO: A Review and Synthesis of the Management Literature. *Journal of Management*, 42(1), 234–268.
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience. *Nature Reviews Neuroscience*, 14, 365–376.
- Chang, Y., Wu, K.-T., & Lin, C.-J. (2024). Board Gender Diversity and Corporate Social Responsibility. *International Journal of Corporate Social Responsibility*, 9, Article 7.
- Chatterji, A. K., Durand, R., Levide, D. I., & Touboul, S. (2016). Do Ratings of Firms Converge? Implications For Managers, Investors and Strategy Researchers. *Strategic Management Journal*, 37(8), 1597–1614.
- Council Regulation (EC) No 2157/2001. (2013). Council Regulation (EC) No 2157/2001 of 8 October 2001 on the Statute for a European Company (SE) as amended on the 13 May 2013 by Council Regulation No 517/2013, in Official Journal L294, 10 November 2001.
- Crace, L., & Gehman, J. (2023). What Really Explains ESG Performance? Disentangling the Asymmetrical Drivers of the Triple Bottom Line. *Organization & Environment*, 36(1), 150–178.
- Crossland, C., & Hambrick, D. C. (2010). Differences in Managerial Discretion Across Countries: How Nation-Level Institutions Affect the Degree to which CEOs Matter. *Strategic Management Journal*, 32(8), 797–819.
- Crossley, R. M., Elmagrhi, M. H., & Ntim, C. G. (2021). Sustainability and Legitimacy Theory: The Case of Sustainable Social and Environ-

- mental Practices of Small and Medium-Sized Enterprises. *Business Strategy and The Environment*, 30(8), 3740–3762.
- Dang, C., Li, Z. F., & Yang, C. (2018). Measuring Firm Size in Empirical Corporate Finance. *Journal of Banking and Finance*, 86, 159–176.
- DCGK C. (2022). Zusammensetzung des Aufsichtsrats.
- Deutscher Bundestag (Ed.) (2021). Drucksache 19/26689. Retrieved October 10, 2024, from <https://dserver.bundestag.de/btd/19/266/1926689.pdf>
- Dienes, D., & Velte, P. (2016). The Impact of Supervisory Board Composition on CSR Reporting. Evidence From the German Two-Tier System. *Sustainability*, 8(1), Article 63.
- DrittelbG. (2021). Drittelbeteiligungsgesetz of 18 May 2004, as amended on 12 August 2021.
- Dube, I., & Jaiswal, N. (2015). Corporate Governance in The Energy Sector. *Jindal Global Law Review*, 6(2), 143–178.
- Eberhardt-Toth, E. (2017). Who Should Be on a Board Corporate Social Responsibility Committee? *Journal of Cleaner Production*, 140, 1926–1935.
- Egon Zehnder (Ed.) (2021). Frischer Wind für mehr Nachhaltigkeit: Die Rolle der Aufsichtsräte. Retrieved October 10, 2024, from <http://www.egonzehnder.com/cdn/serve/article-pdf/1618844868-28750033da3c9bf35bbc47abba9dc393.pdf>
- Elkington, J. (2004). Enter the Triple Bottom Line (A. Henriques & J. Richardson, Eds.), 1–16.
- Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47(2), 427–465.
- Farri, E., Cervini, P., & Rosani, G. (2022). How Sustainability Efforts Fall Apart. Retrieved October 10, 2024, from <https://hbr.org/2022/09/how-sustainability-efforts-fall-apart>
- Flammer, C., Hong, B., & Minor, D. (2019). Corporate Governance and the Rise of Integrating Corporate Social Responsibility Criteria in Executive Compensation: Effectiveness and Implications For Firm Outcomes. *Strategic Management Journal*, 40(7), 1097–1122.
- Frauen in die Aufsichtsräte e.V. (Ed.) (2020). Women-on-Board-Index 185. Retrieved October 10, 2024, from <https://www.wob-index.de/wobmedia/documents/wob185/2020/epaper/>
- Frauen in die Aufsichtsräte e.V. (Ed.) (2022). Women-on-Board-Index 185. Retrieved October 10, 2024, from <https://www.yumpu.com/de/document/read/67040096/women-on-board-index-185>
- Frauen in die Aufsichtsräte e.V. (Ed.) (2024). Women-on-Board-Index 185. Retrieved October 10, 2024, from <https://www.yumpu.com/de/document/read/68757577/wob-index-185-studie-ranking-2024>
- Gallo, P. J., & Christensen, L. J. (2011). Firm Size Matters: An Empirical Investigation of Organizational Size and Ownership on Sustainability-Related Behaviors. *Business & Society*, 50(2), 315–349.
- Gartner (Ed.) (2021). The ESG Imperative: 7 Factors for Finance Leaders to Consider. Retrieved October 10, 2024, from <https://www.gartner.com/smarterwithgartner/the-esg-imperative-7-factors-for-finance-leaders-to-consider>
- Gerwing, T., Kajüter, P., & Wirth, M. (2022). The Role of Sustainable Corporate Governance in Mandatory Sustainability Reporting Quality. *Journal of Business Economics*, 92(3), 517–555.
- Glass, C., Cook, A., & Ingersoll, A. R. (2016). Do Women Leaders Promote Sustainability? Analyzing the Effect of Corporate Governance Composition on Environmental Performance. *Business Strategy and the Environment*, 25(7), 495–511.
- Goodstein, J., Gautam, K., & Boeker, W. (1994). The Effects of Board Size and Diversity on Strategic Change. *Strategic Management Journal*, 15(3), 241–250.
- Hambrick, D. C., & Mason, P. A. (1984). Upper Echolons: The Organization as a Reflection of its Top Managers. *The Academy of Management Review*, 9(2), 193–206.
- Haque, F., & Jones, M. J. (2020). European Firms' Corporate Biodiversity Disclosures and Board Gender Diversity from 2002 to 2016. *The British Accounting Review*, 52(2), Article 100893.
- Harjoto, M., Laksmana, I., & Lee, R. (2015). Board Diversity and Corporate Social Responsibility. *Journal of Business Ethics*, 132(4), 641–660.
- HGB. (2024). Handelsgesetzbuch of 10 May 1897, as amended on 17 April 2024.
- Hill, C. W., & Jones, T. (1992). Stakeholder-Agency Theory. *Journal of Management Studies*, 29(2), 131–154.
- Hong, B., Li, Z. F., & Minor, D. (2016). Corporate Governance and Executive Compensation for Corporate Social Responsibility. *Journal of Business Ethics*, 136(1), 199–213.
- Horbach, J., & Jacob, J. (2018). The Relevance of Personal Characteristics and Gender Diversity for (Eco-)Innovation Activities at the Firm-Level: Results From a Linked Employer-Employee Database in Germany. *Business Strategy and the Environment*, 27(7), 924–934.
- Hrasky, S. (2011). Carbon Footprints and Legitimation Strategies: Symbolism or Action? *Accounting, Auditing & Accountability Journal*, 25(1), 174–198.
- Huther, J. (1997). An Empirical Test of the Effect of Board Size on Firm Efficiency. *Economics Letters*, 54, 259–264.
- International Organization for Standardization (Ed.) (n.d.). ISO 26000:2010(en). Retrieved October 10, 2024, from <https://www.iso.org/obp/ui/en/#iso:std:iso:26000:ed-1:v1:en>
- Isidro, H., & Sobral, M. (2015). The Effects of Women on Corporate Boards on Firm Value, Performance, and Ethical and Social Compliance. *Journal of Business Ethics*, 132(1), 1–19.
- Jensen, M., & Meckling, W. (1976). The Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics*, 3(4), 305–360.
- Jizi, M. (2017). The Influence of Board Composition on Sustainable Development Disclosure. *Business Strategy and the Environment*, 26(5), 640–655.
- Kanter, R. M. (1977). Some Effects of Proportions on Group Life. *American Journal of Sociology*, 82(5), 965–990.
- Khaled, R., Ali, H., & Mohamed, E. K. (2021). The Sustainable Development Goals and Corporate Sustainability Performance: Mapping, Extent and Determinants. *Journal of Cleaner Production*, 311, Article 127599.
- Kidwell, J. M., Stevens, R. E., & Bethke, A. L. (1987). Differences in Ethical Perceptions Between Male and Female Managers: Myth or Reality? *Journal of Business Ethics*, 6(6), 489–493.
- Kim, S. J., Kim, H., & Atukeren, E. (2023). Effects of Board Independence on Greenhouse Gas Emissions and Financial Consequences: Evidence from South Korea. *Environments*, 10(3), Article 56.
- Kirsch, A., & Wrohlich, K. (2020). More Women on Supervisory Boards: Increasing Indications that the Effect of the Gender Quota Extends to Executive Boards. *DIW Weekly Report*, (4+5), 44–49.
- Konrad, A. M., Kramer, V., & Erkut, S. (2008). Critical Mass: The Impact of Three or More Women on Corporate Boards. *Organizational Dynamics*, 37(2), 145–164.
- Liu, C. (2018). Are Women Greener? Corporate Gender Diversity and Environmental Violations. *Journal of Corporate Finance*, 52, 118–142.
- Loukil, N., Yousfi, O., & Yerbanga, R. (2019). Does Gender Diversity on Boards Influence Stock Market Liquidity? Empirical Evidence From The French Market. *Corporate Governance*, 19(4), 669–703.
- Mattingly, J. E., & Berman, S. L. (2006). Measurement of Corporate Social Action: Discovering Taxonomy in the Kinder Lydenburg Domini Ratings Data. *Business & Society*, 45(1), 20–46.
- Menicucci, E., & Paolucci, G. (2022). Board Diversity and ESG Performance: Evidence from the Italian Banking Sector. *Sustainability*, 14(20), Article 13447.
- MitbestG. (2021). Mitbestimmungsgesetz of 01 July 1976, as amended on 07 August 2021.
- Neter, J., Wasserman, W., & Kutner, M. H. (1983). *Applied Linear Regression Models*.
- Nuber, C., & Velte, P. (2021). Board Gender Diversity and Carbon Emissions: European Evidence on Curvilinear Relationships and Critical Mass. *Business Strategy and the Environment*, 30(4), 1958–1992.
- Pierce, J. R., & Aguinis, H. (2013). The Too-Much-of-a-Good-Thing Effect in Management. *Journal of Management*, 39(2), 313–338.
- Poole, M. A., & O'Farrell, P. N. (1971). The Assumptions of the Linear Regression Model. *Transactions of the Institute of British Geographers*, (52), 145–158.

- Rao, K. K., & Tilt, C. (2021). Gender and CSR Decisions: Perspectives from Australian Boards. *Meditari Accountancy Research*, 29(1), 60–85.
- Refinitiv (Ed.) (2022). Environmental, Social and Governance Scores from Refinitiv. Retrieved October 10, 2024, from https://www.lseg.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf
- Schwartz-Ziv, M. (2017). Gender and Board Activeness: The Role of a Critical Mass. *Journal of Financial and Quantitative Analysis*, 52(2), 751–780.
- Shahbaz, M., Karaman, A. S., Kilic, M., & Uyar, A. (2020). Board Attributes, CSR Engagement, and Corporate Performance: What is the Nexus in the Energy Sector? *Energy Policy*, 143, Article 111582.
- Srinidhi, B., Gul, F. A., & Tsui, J. (2011). Female Directors and Earnings Quality. *Contemporary Accounting Research*, 28(5), 1610–1644.
- Steger, T., & Jahn, S. (2019). Roles and Tasks of German Supervisory Boards: An Explanatory View (J. Gabriellson, W. Khlif, & S. Yamak, Eds.), 358–378.
- STOXX (Ed.) (2024). DAX Equity Index Methodology Guide. Retrieved October 10, 2024, from https://www.stoxx.com/document/Indices/Common/Indexguide/DAX_Equity_Index_Methodology_Guide.pdf
- Sumarta, N. H., Prabowo, M. A., Amidjaya, P. G., Supriyono, E., & Prameswar, A. P. (2021). CEO Characteristics and Environmental Performance: Evidence from Indonesian Banks. *International Journal of Business and Society*, 22(2), 1015–1033.
- Tagliatela, J., Maffiola, K. P., Barontini, R., & Testa, F. (2023). Board of Directors' Characteristics and Environmental SSGs Adoption: An International Study. *Corporate Social Responsibility and Environmental Management*, 30(5), 2490–2506.
- United Nations. (1987). *Our Common Future: Report of The World Commission on Environment and Development*. Retrieved October 10, 2024, from <http://www.un-documents.net/ocf-ov.htm>
- Vafeas, N. (1999). Board Meeting Frequency and Firm Performance. *Journal of Financial Economics*, 53(1), 113–142.
- Velte, P. (2016a). Sustainable Management Compensation and ESG Performance – The German Case. *Problems and Perspectives In Management*, 14(4), 17–24.
- Velte, P. (2016b). Women on Management Board and ESG Performance. *Journal of Global Responsibility*, 7(1), 98–109.
- Velte, P. (2020). Do CEO Incentives and Characteristics Influence Corporate Social Responsibility (CSR) and Vice Versa? A Literature Review. *Social Responsibility Journal*, 16(8), 1293–1323.
- Velte, P. (2022). Does Sustainable Board Governance Drive Corporate Social Responsibility? A Structured Literature Review on European Archival Research. *Journal of Global Responsibility*, 14(2), 1–43.
- Velte, P. (2023). Sustainable institutional investors and corporate biodiversity disclosure: Does sustainable board governance matter? *Corporate Social Responsibility and Environmental Management*, 30, 3063–3074.
- Visentin, D. C., Cleary, M., & Hunt, G. E. (2020). The Earnestness of Being Important: Reporting Non-Significant Statistical Results. *Journal of Advanced Nursing*, 76(4), 917–919.
- Waddock, S. A., & Graves, S. B. (1997). The Corporate Social Performance-Financial Performance Link. *Strategic Management Journal*, 18(4), 303–319.
- Williams, R. J. (2003). Women on Corporate Boards of Directors and their Influence on Corporate Philanthropy. *Journal of Business Ethics*, 42(1), 1–10.
- Yarram, S. R., & Adapa, S. (2021). Board Gender Diversity and Corporate Social Responsibility: Is There a Case for Critical Mass? *Journal of Cleaner Production*, 278, Article 123319.
- Zahra, S. A., & Pearce II, J. A. (1989). Boards of Directors and Corporate Financial Performance: A Review and Integrative Model. *Journal of Management*, 15(2), 297–334.
- Zaid, M. A., Abuhijleh, S. T., & Pucheta-Martinez, M. C. (2020). Ownership Structure, Stakeholder Engagement, and Corporate Social Responsibility Policies: The Moderating Effect of Board Independence. *Corporate Social Responsibility and Environmental Management*, 27(3), 1344–1360.
- Zhang, L. (2012). Board Demographic Diversity, Independence, and Corporate Social Performance. *Corporate Governance*, 12(5), 686–700.
- Zimek, M., & Baumgartner, R. (2017). Corporate Sustainability Activities and Sustainability Performance of First and Second Order.



Implicit Measurement of the Moral Self-Image Using the Go/No-Go Association Task (GNAT) - An Empirical Investigation of the Convergent Validity Between Explicit and Implicit Measures

Louisa Felicitas Bläßer

Technical University of Munich

Abstract

While people are increasingly aware of climate change, many still resist lifestyle changes. Research now focuses on understanding conscious (explicit) and unconscious (implicit) attitudes to encourage sustainable behavior. This thesis used the Go/No-Go Association Task (GNAT) to measure participants' implicit moral self-image and examine its correlation with an explicit moral self-image questionnaire, indicating convergent validity and effective application of the GNAT as an implicit measure of the moral self-image. After applying exclusion criteria, 68 participants were randomly assigned to two groups with different word lists. Results showed that repeated exposure to fewer words in group A led to little or no correlation, while group B, using more varied words, showed higher correlation and good convergent validity. This demonstrates that the GNAT effectively measures moral self-image when learning effects are avoided. The findings offer insights into implicit attitudes that influence decisions and yield practical implications for different stakeholders. This thesis contributes through its experimental design, adapted exclusion criteria, and sample correction of all perfect responses, validating the GNAT as an implicit measure and offering a foundation for future research.

Keywords: convergent validity; explicit measures; go/no-go association task (GNAT); implicit measures; moral self-image

1. Introduction

Have you ever ordered something online, used non-recyclable packaging, chosen a non-organic product, not separated food waste appropriately, or traveled by plane? The answer is likely yes, as we face many sustainable decisions daily. Unsustainable behavior is perceived as immoral, as people are increasingly aware of their impact on climate change (Sachdeva et al., 2015). But why do people engage in unsustainable and, consequently, immoral behaviors?

There has been an increasing focus on understanding the psychological drivers that motivate immoral behavior in the last decades, especially with the intensifying global climate crisis (Sachdeva et al., 2015). The ultimate goal is to use this knowledge to nudge people further into being more sustainable (Fischer et al., 2012; Sachdeva et al., 2015). Therefore, it is essential to understand why people behave immorally and how people's morality can be measured.

Traditional explicit measures often fall short of capturing the perception of people's moral selves. Social desirability biases influence the answers given in such self-report questionnaires (Crowne & Marlowe, 1960). In recent decades, various implicit measures have been developed to measure unconscious attitudes. The most famous method is the Implicit Association Test (IAT), which was further developed into the Go/No-Go Association Task (GNAT). Previous research has already implicitly assessed the moral self-image (perception

I would like to thank Konrad Kober, my Bachelor's thesis supervisor, for his dedicated support, continuous encouragement, and constructive feedback throughout the duration of this thesis. I am especially grateful for his valuable assistance and expertise in the experimental design. Additionally, I would like to thank Prof. Dr. Alwine Mohnen from the Chair of Corporate Management at the Technical University of Munich for serving as my examiner.

of one's own morality) with the IAT, yielding promising results (Perugini & Leone, 2009). However, little research has been devoted to the GNAT, and it has only been used once to capture the moral self-image (Ferguson, 2018).

This bachelor's thesis uses the GNAT to measure the implicit moral self-image and to examine the convergent validity (the ability of two measures to capture a joint construct (Carlson & Herdman, 2012)) of this implicit measurement method by correlating it with an explicit moral self-image questionnaire. It aims to answer the following research question: "To what extent can the Go/No-Go Association Task (GNAT) be effectively applied to measure moral self-image, and is there a correlation between the outcomes of this method and the explicit moral self-image?"

The timing for this research is crucial, as it aligns with the growing interest in sustainable practices and the need to deepen the understanding of the internal, implicit motivations behind a behavior change (Mazar & Zhong, 2010; Sachdeva et al., 2015; Schlegelmilch & Simbrunner, 2019). Applying implicit psychological tests, such as the GNAT, to consumer behavior or marketing strategies opens up a new field of research. These measures improve the assessment of implicit attitudes toward products or brands because the implicit test procedures are based on less biased, unconscious answers and reactions and, therefore, are very valuable for subsequent analyses. An effective and validated tool to determine people's moral self-image could provide essential insights for different stakeholders to influence consumers toward more sustainable and moral choices.

This thesis is structured as follows:

Chapter 1 briefly introduces the topic and aim of this thesis. Chapter 2 reviews current research literature, presenting theories of moral behavior, important definitions, and measurement methods, leading into Chapter 3, which covers the methodology and detailed research design of the performed GNAT experiment. Chapters 4 and 5 present the compelling results and elaborate on the experiment's key findings, discussing its limitations, suggestions for future research, and implications for practitioners in management and policy. Chapter 6 provides a comprehensive summary of the main findings of the experiment, reflecting on the research's significance.

2. Theoretical Background

A wide range of theories are trying to explain why people behave immorally. The rational economic model expects people to behave immorally whenever their potential gain exceeds their expected punishment since it is the best choice economically (Becker, 1968). Following this reasoning, people should behave immorally every time they could potentially gain more than they would lose. In contrast, it can be observed that people intrinsically limit their immorality and avoid too much lying if it threatens their perception of their own morality (moral self-image) (Mazar et al., 2008; Sachdeva et al., 2009). It appears that the idea of an entirely rational person (e.g., homo economicus (Melé & Can-

tón, 2014)) does not apply to most people and situations. Instead, an internal force seems to restrict people from exploiting the potential benefits of cheating to its full extent (Cornelissen et al., 2013; Mazar et al., 2008).

It becomes evident that people face an internal conflict whenever they have an opportunity to cheat (Barkan et al., 2015; Mazar et al., 2008). This ethical dissonance is a state of tension that occurs when people are either tempted to benefit from their immoral behavior or to uphold a positive moral self-image, also known as moral self-concept¹ (Mazar et al., 2008). Festinger (1957) describes this state as cognitive dissonance and argues that its presence motivates people to subsequent action, which reduces this dissonance. People developed different strategies to engage in immoral behavior to resolve this internal conflict and distressing state, particularly without updating (and depressing) their moral self-image.

Erikson (1964) explains this motivating force as the intrinsic need for people to act according to their (moral) identity. Researchers interpret moral identity, defined as "the use of moral values to define the self" (Johnston et al., 2013, p. 209), as a moderator and motivational driver to act morally (Aquino & Reed, 2002; Blasi, 1993; Erikson, 1964).

The following section will introduce different theories that investigate why people behave immorally.

2.1. Moral Theories

2.1.1. Self-Concept Maintenance

Mazar et al. (2008) propose a theory of self-concept maintenance. They argue that people try to balance maintaining an honest self-concept and gaining from lying. According to Mazar et al. (2008), people would cheat to a certain extent as long as they do not need to update their moral self-concept (of being honest). This compromise allows them to benefit from cheating without negatively impacting their moral self-image. The authors also suggest that people use different techniques to decide on this motivational dilemma and to determine the degree to which cheating aligns with their moral self.

A powerful technique is, for instance, self-serving justification (Shalvi et al., 2015). It suggests that people would try to find reasons for questionable behavior to make it seem less immoral when their moral self-image is threatened. Shalvi et al. (2015) distinguish between pre-violation justification (before the immoral action) and post-violation justification (after the immoral action).

Pre-violation justification excuses the immoral action and thus reduces the threat to the moral self-concept beforehand (Shalvi et al., 2015). There are several strategies for this pre-violation justification. Examples are ambiguous actions, altruistic cheating, and moral licensing (Shalvi et al., 2015). Whenever the norms and rules for a situation are ambiguous, the actor could invent facts and reasons to justify his²

¹ These terms can be used interchangeably (Jordan et al., 2015)

² For better readability, only the pronouns "he/him/his" are used throughout this thesis

actions. If a lie does not cause harm to other people but instead would help to benefit the actor and other people, it is more likely to observe cheating (altruistic cheating) (Erat & Gneezy, 2012). Through moral licensing, people justify their bad behavior with their initial good actions (Merritt et al., 2010; Shalvi et al., 2015). In contrast, post-violation justification is a tool to justify immoral behavior after the action has already been conducted. This could be, for instance, through (partially) confessing or distancing themselves from their action by looking at others' immoral behavior (Shalvi et al., 2015).

People generally try to maintain or even enhance their moral self-image (Jordan et al., 2015; Mazar et al., 2008; Shalvi et al., 2015). They do this by behaving morally or biasing their cognitive perception with examples like self-serving justifications (Monin & Jordan, 2009). Monin and Jordan (2009) argue that people who value morality greatly pay more attention to their moral self-image, and deviations from their moral self-concept impact their self-worth more significantly compared to people with a lower importance on being moral. This moral self-image can also be influenced by previous and current situations. A deviation from their aspired level motivates people to take subsequent actions to reduce this dissonance (Cornelissen et al., 2013; Jordan et al., 2015). Monin and Jordan (2009) refer to that as behavior-generating power. To predict peoples' behavior, their individual moral self-image, which fluctuates and deviates over time (Jordan et al., 2015), must be considered.

2.1.2. Moral Balancing Model

There are two contrasting approaches when predicting peoples' moral actions after they have acted morally or immorally. Either the actor behaves consistently with his initial action, or the subsequent behavior is the opposite of his previous action (moral balancing).

Freedman and Fraser (1966) introduced the Foot-In-The-Door-Technique, which is nowadays widely used in negotiation strategies and a great example of consistent behavior. They elaborate that people who already agreed to do a small favor were more likely to agree to do a second, even larger favor.

An example of consistent moral behavior would be if a man returns a lost wallet to the owner after offering a seat in public transport to an elderly woman. The negative case, which still reflects consistent behavior, would be that the man does not offer his seat to the elderly woman and keeps the wallet as well.

Mullen and Monin (2016) argue that people show consistent behavior when they focus abstractly on values and their initial behavior. In contrast, people exhibit a balancing behavior when they think more concretely about their initial behavior and what they have accomplished with it. This alternating pattern is described as moral balancing. After a previous immoral action, the actor behaves morally in the subsequent action, or vice versa.

This moral balancing model was developed in 1990 by the psychologist Mordecai Nisan. It states that people con-

sider previous behavior when making moral decisions. According to Nisan (1990), people try to balance their current moral self around a fixed personal moral standard (equilibrium). This personal reference point is essential for people as they constantly compare their current state with this self-set standard, which they want to maintain over time (Miller & Effron, 2010). Nisan (1990) assumes that when their moral status drops below a personal tolerable level, people will refrain from doing an immoral action. However, this satisfactory level of morality is lower than the ideal level, and those minimum requirements are determined mainly by a person's moral identity (Nisan, 1990).

Following this reasoning, a person who recently did something immoral would instead choose an altruistic action in order to compensate for the previously generated deficit in his own moral balance (moral cleansing). A person who is currently in moral surplus would be more likely to perform a subsequent selfish action (moral licensing) (Nisan, 1990). In other words, balancing happens when a moral initial behavior leads to the opposite in a subsequent behavior (Jordan et al., 2011; Mullen & Monin, 2016; Zhong et al., 2010).

When balancing a previous action, these two directions can be observed: moral cleansing and moral licensing.

Moral Cleansing

Moral cleansing (or moral compensation) happens when a previous immoral behavior causes a subsequent moral behavior (Mullen & Monin, 2016; Perkins et al., 2024). This can be explained by an analogy of a moral bank account, the moral credits model (Perkins et al., 2024). If a person's metaphorical moral bank account is in deficit, he wants to rebalance it with a subsequent moral behavior (Nisan, 1990). This could be done by performing a morally good action or refraining from immoral actions, such as cheating (Cornelissen et al., 2013). Continuing with the previous example, a man who did not offer his seat on the bus to an elderly woman would be more likely to return a lost wallet to its owner to compensate for this deficit in his moral balance. Researchers explain this effect through people's motivation and willingness to invest effort to repair their shortfalls (Jacobsen et al., 2018).

Additionally, there is strong evidence that people need to physically cleanse themselves after behaving immorally. Zhong and Liljenquist (2006) show that people who recall an immoral act would be more likely to choose antiseptic wipes compared to other products. They explain that the participants need to wash away their sins and cleanse themselves after their moral purity has been threatened.

Moral Licensing

The moral licensing effect describes the contrasting and somewhat counterintuitive observation: Good previous behavior leads to less positive or even bad behavior. In other words, people justify their bad behavior with their previous

good action (Jacobsen et al., 2018; Merritt et al., 2010).

Moral licensing can be explained from two different perspectives: the moral credits model and the moral credentials model.

The moral credits model explains the licensing effect as people accumulate credits in their hypothetical moral bank account when they do something good. They can use these credits and “withdraw” them to justify subsequent negative behavior while maintaining an overall positive balance (Efron & Monin, 2010; Merritt et al., 2010; Miller & Efron, 2010). Moral licensing starts with a surplus in the moral bank account and withdraws credits to allow people to perform a negative action (Perkins et al., 2024).

The second explanation, the moral credentials model, explains the moral licensing effect with a different interpretation of the subsequent behavior. According to Monin and Miller (2001), people are less likely to interpret their subsequent behavior as immoral after they have performed an initial moral act. Instead of earning a right to perform this immoral act without punishment, the initial moral behavior has provided a lens through which the following behavior is interpreted differently (Mullen & Monin, 2016). This process is more likely when the subsequent behavior is ambiguous and can be interpreted positively (Mullen & Monin, 2016). For example, by recommending a woman for one job, people built positive credentials as being someone without prejudice and were more willing to express that a man was better suited for a second job (Monin & Miller, 2001). In this experiment by Monin and Miller (2001), the second behavior was ambiguous. It could be explained by illegitimate or legitimate motives (sexism or pragmatism). The credentials of not being sexist, e.g., established through actively recommending a woman for the first job, help to interpret the second action positively, e.g., favoring a man for the second job, due to the actor’s history, without affecting the actor’s moral self-image (Monin & Jordan, 2009; Monin & Miller, 2001).

Both models explain that a previous moral action can lead to immoral or questionable behavior later on. The key difference is that in the moral credits model, the actor is fully aware of the second action’s immorality but decides to afford this decrease in his overall moral balance. In contrast, in the moral credentials model, the positive first action helps to disambiguate and interpret the second action differently (Monin & Jordan, 2009). To clarify this tension between the two models, Monin and Jordan (2009) suggest that the moral credits model is at work in unambiguous cases, where the meaning of the target behavior is clearly interpreted as immoral and unaffected by the previous action, while the moral credentials apply for ambiguous cases. However, both models predict the same behavior and support the importance of acknowledging a dynamic moral self-image. These models suggest that recent actions shape a person’s moral self-image and influence his future moral behavior (Monin & Jordan, 2009).

Moral Self-Image in the Moral Balancing Model

The moral credits model describes a mechanism for people to balance their moral or immoral behavior with an accumulated or depleted moral bank account, reflecting the increase or decrease of the moral self-image, respectively (Merritt et al., 2010; Zhong et al., 2010). This enables people to repair their moral self-image by compensating for their selfish actions afterward (moral cleansing) (Perkins et al., 2024; Schlegelmilch & Simbrunner, 2019) or using their bolstered moral self-image (from a previous action) to perform a subsequent immoral act (moral licensing) (Cornelissen et al., 2013; Efron & Monin, 2010; Monin & Jordan, 2009; Nisan, 1990).

This emphasizes that the moral self-image plays a central role in moral decision-making. Its discrepancies from the actor’s personal standard (equilibrium) motivate balancing behavior (Nisan, 1990). While the balancing could be observed, and there is empirical evidence (Cornelissen et al., 2013; Lee & Hsieh, 2013; Ploner & Regner, 2013), measuring the moral self-image is also important. Cornelissen et al. (2013) first attempted to measure the moral self-image with a scale of differences between the desired and the perceived moral self. Jordan et al. (2015) developed this scale further to provide a tool that actively and explicitly measures the moral self-image. However, there is still little empirical evidence of the deviations and fluctuations in time of the moral self-image (Perkins et al., 2024).

2.2. Definitions

As the previous section illustrated, different theories try to explain immoral behavior. Given the important role of the moral self, moral psychology increasingly shifted its focus to it to extend moral reasoning and predict behavior (Monin & Jordan, 2009). Before introducing some measurement methods, two terms must be defined accordingly in the context of the moral self: moral identity and moral self-image.

2.2.1. Moral Identity

Aquino and Reed (2002) define moral identity “as a self-conception organized around a set of moral traits.” (p. 1424). They suggest that moral identity is relatively stable over time and identify two dimensions: Internalization and Symbolization. Internalization describes how important it is for a person to have (nine) moral traits: “caring, compassionate, fair, friendly, generous, helpful, hardworking, honest, and kind.” (Aquino & Reed, 2002, p. 1426). Symbolization describes the degree to which a person wants to be seen as moral or demonstrate these traits through their actions to others (Aquino & Reed, 2002). The researchers propose that people behave morally when they assess a specific moral trait as essential for their self-concept. Moral identity should, therefore, be a motivational driver for acting consistently (Aquino & Reed, 2002) and is the basis for moral motivation (Erikson, 1964; Nisan, 1990).

To measure moral identity actively, Aquino and Reed (2002) asked participants to rate how important it is for

them to possess these traits (Internalization) and if they participate in activities (e.g., hobbies), wear clothes or buy products that identify them as having these characteristics (Symbolization) (Aquino & Reed, 2002).

2.2.2. Moral Self-Image

Jordan et al. (2015) introduced the concept of the moral self-image to explain how the self-perception of the individual's morality fluctuates. They define the moral self-image as the malleable and dynamic moral self-concept.

A person's moral self-image can be described as the answer to the question "How moral am I?" (Monin & Jordan, 2009, p. 347). This reflects exactly how morally individuals see themselves at any point in time. The moral self-image is part of the dynamic working self-concept, the malleable part of the self (Jordan et al., 2015). It is completely subjective and only measures how moral persons perceive themselves (Jordan et al., 2015). Monin and Jordan (2009) highlight that individuals can constantly show differences in their moral self-image, as it can be lowered or bolstered through previous actions, which motivates subsequent behavior. The researchers agreed that the moral self-image has a behavior-generating power (Jordan et al., 2015; Monin & Jordan, 2009).

Due to the lack of previous empirical measurement methods, Jordan et al. (2015) introduced an explicit nine-point Likert scale to measure the moral self-image as highly connected to the traits of a typical moral person (based on the traits introduced by Aquino and Reed (2002)). This elaborated scale has been used as an explicit moral self-image measure in previous research (Ferguson, 2018) to investigate the convergent validity, which "reflects the extent to which two measures capture a common construct." (Carlson & Herdman, 2012, p. 18).

2.3. Measurement Methods

2.3.1. Explicit vs. Implicit Measures

In order to understand, predict, and control human behavior, psychologists have been trying to measure people's cognitive processes, attitudes, and self-image (de Houwer, 2006).

A straightforward approach is to conduct a survey and actively ask participants about their opinions toward a situation or an object. This explicit method is easy to conduct, comprehensible, and easily measured (de Houwer, 2006). The most common approach for measuring the moral identity or the moral self-image is letting participants rate different personality traits on a Likert scale. This approach assesses how important these personality traits are for them (moral identity) (Aquino & Reed, 2002) or how much they are already fulfilling some characteristics compared to the person they want to be (moral self-image) (Jordan et al., 2015). However, despite their wide use (Asendorpf et al., 2002), these surveys might be subject to impression management (Paulhus, 1984), which means that participants include answers to be seen in a favorable light. When being asked, people are

influenced by concerns about their self-presentation (Doherty & Schlenker, 1991; Schnabel et al., 2007) and social desirability (Crowne & Marlowe, 1960), which could incentivize them to give socially conform answers to the interviewer. Additionally, these surveys are limited to the introspective personality and might not reflect a person's entire personality (Schnabel et al., 2007).

Because of these disadvantages, new implicit measures have been developed. Initially introduced in social psychology, implicit measures are now widely applied across different disciplines and commonly used in psychology (de Houwer et al., 2009). But what is an implicit measure exactly? de Houwer (2006) suggests using the synonym automatic when explaining implicit effects. A process can be called automatic when it still operates, although the participants are unaware of results, stimulus, or procedure, do not have a specific goal, or do not invest many cognitive resources (de Houwer, 2006). Following this argumentation, the same should apply to an implicit measure. This measure intends to get an immediate (automatic) response from people without them being aware of it or involving their cognitive thinking. Such an implicit or indirect measurement method could be used to measure a person's unconscious attitude (Bartels & Schoenrade, 2022; Schnabel et al., 2007).

2.3.2. Implicit Association Test (IAT)

As a way to avoid biases of explicit measurement, Greenwald et al. developed the Implicit Association Test (IAT) in 1998. This test aims to measure the relative implicit association strength of two contrasting concepts (target categories) (e.g., FLOWER-INSECT) and PLEASANT-UNPLEASANT³ (evaluation attribute) (Greenwald et al., 1998).

In their initial experiment, all participants should react by pressing an assigned key on the left or right. In different blocks, a target category and an attribute are assigned to one key. For example, the left key is assigned to FLOWER + PLEASANT, whereas the right key is assigned to INSECT + UNPLEASANT. Whenever a stimulus (either a FLOWER, an INSECT, a PLEASANT, or an UNPLEASANT word) appears on the screen, the participant should press the assigned key (Greenwald et al., 1998). In other words, the stimuli should be classified into four mutually exclusive categories (FLOWER, INSECT, PLEASANT, or UNPLEASANT) (Greenwald et al., 1998; Schimmack, 2021). Participants are required to distinguish between words referring to INSECT + FLOWER and words referring to PLEASANT + UNPLEASANT words. For instance, with the assigned keys described above, the stimuli tulip or happy should be assigned to the left key (FLOWER + PLEASANT), whereas wasp or rotten should be assigned to the right key (INSECT + UNPLEASANT) (Greenwald et al., 1998). After the first combined task of discrimination between the target categories and evaluation attributes, a second block with a reversed combined

³ The categories (target categories and evaluative attributes) are written in capital letters.

task was conducted. In this reversed combined task, one key was assigned to INSECT + PLEASANT, and the other key was assigned to FLOWER + UNPLEASANT.

It is fundamentally assumed that participants' response time is faster when the association between the target category and the evaluative attribute is stronger (de Houwer, 2006; de Houwer, 2001; Greenwald et al., 1998; Johnston et al., 2013). This means that the pairing FLOWER + PLEASANT should be easier compared to the INSECT + PLEASANT pairing if the association between FLOWER and PLEASANT words is stronger (de Houwer, 2006; de Houwer, 2001; Greenwald et al., 1998; Johnston et al., 2013). With this experiment, Greenwald et al. (1998) provided significant results demonstrating that the incompatible combination of INSECT + PLEASANT was more challenging to confirm, and participants had longer response times compared to the compatible combination of FLOWER + PLEASANT. The authors explain this effect with a stronger association and familiarity between FLOWER + PLEASANT words and than between INSECT + PLEASANT words, indicating a more positive attitude toward FLOWERS than INSECTS (Greenwald et al., 1998).

While the IAT was quite revolutionary, set new standards, and offered new opportunities, criticism about the IAT and implicit measures, in general, needs to be addressed. There are concerns about its construct validity (Schimmack, 2021), its capability to predict behavior (Bartels & Schoenrade, 2022; Brownstein et al., 2020), and its temporal instability (Brownstein et al., 2020; Schimmack, 2021).

Schimmack (2021) raises concerns that there is no consensus about what the IAT measures and that it is difficult to compare if it measures something different than explicit measures. This problem has been recognized by the dual attitudes model (Wilson et al., 2000) (also known as the double dissociation model (Perugini, 2005)), which clearly distinguishes implicit and explicit attitudes into two systems (Wilson et al., 2000). According to this model, implicit measures predict impulsive, spontaneous, and automatic behavior, while explicit measures predict controlled and conscious behavior (Wilson et al., 2000). In agreement with the double dissociation model, Johnston et al. (2013) suggest that implicit and explicit attitudes can only be measured with implicit or explicit measurement methods, respectively.

The contrasting perspective describes an additive view, where both types of attitude describe a "different portion of variance in the same criterion" (Perugini, 2005, p. 29). Fazio and Olson (2003) argue that both measures assess the same construct and explain a potential difference between the measurement methods with participants' deliberative control strategies.

The IAT and other implicit measurement methods can add predictive insights to self-report measures (Brownstein et al., 2020) and investigate the implicit moral self-image. Considering the concerns about the behavior predictability of the IAT, there are several studies about predictions of voting behavior with the IAT. For example, Friese et al. (2007) successfully predicted the voting behavior and attitudes for the

German parliamentary elections in 2002. They used a single-target IAT, where one key was assigned to an evaluative attribute and the target category, while the second key was assigned only to the opposing attribute. This single-target IAT yielded excellent validity in predicting voting behavior (Friese et al., 2007).

2.3.3. Go/No-Go Association Task (GNAT)

To expand the use of implicit measurement methods and the Implicit Association Test (IAT), a new method was developed by Nosek and Banaji (2001). They introduced the Go/No-Go Association Task (GNAT), which mainly focuses on the error rate as the dependent variable to measure the strength of implicit associations (Greenwald et al., 1998; Nosek & Banaji, 2001).

Unlike the previously known implicit methods, in the GNAT, only a single concept (target category / e.g., ME) is evaluated considering one attribute dimension (evaluative attribute / e.g., GOOD) (Bassett & Dabbs, 2005; Ferguson, 2018; Nosek & Banaji, 2001). The GNAT does not need two contrasting concepts (two target categories); hence, it is more flexible and can reveal new aspects of social cognition. Another difference is that only one response (key) is required for the GNAT (Nosek & Banaji, 2001), simplifying the experimental setup.

During the task, a target stimulus (signal item) or a distracter stimulus (noise item) is presented on the screen for some milliseconds. Following the experimental design and example of Ferguson (2018) for the GNAT, when a stimulus word that is similar to either the attribute (e.g., GOOD) or the target category (e.g., ME) is shown, the participant should press the space bar (or any key) to give a Go response. On the contrary, when the word on the center of the screen does not match the attribute or the target category (= distractor), no response (No-Go) is required, and the participant should refrain from pressing any key.

According to Nosek and Banaji (2001), the strength of association in the GNAT is determined by how well stimuli words associated with the target category and the attribute (for example, ME + GOOD) are distinguished from distractor items unrelated to these concepts. The authors suggest that the sensitivity between the pairing conditions (in this example, ME + GOOD or ME + BAD) illustrates the strength of the association between the target category and the evaluative attribute (Greenwald et al., 1998; Nosek & Banaji, 2001). In general, the faster and/or the fewer errors (and therefore easier) the response, the stronger the association. Greenwald et al. (1998) and Nosek and Banaji (2001) argue that both error rates and average response times can provide information about task performance due to a speed-accuracy trade-off. Nevertheless, most implicit measures focus solely on response times "as the dependent variable and therefore may lose relevant information contained in error rates." (Nosek & Banaji, 2001, p. 628).

Previous research in psychology has used the GNAT to investigate different implicit attitudes. For example, implicit spider fear associations (Teachman, 2007), implicit bias in

phrasing drug addiction (Ashford et al., 2019), and implicit attractiveness beliefs of people who are constantly worrying about their physical appearance (Buhlmann et al., 2011). Those studies provide significant insights into the reliability and validity of the GNAT. In those applications, the researchers suggest that the GNAT is an effective tool for measuring involuntary associations and might help measure implicit associations, especially since it does not require a comparison category on a second key (Buhlmann et al., 2011; Teachman, 2007; Williams & Kaufmann, 2012). Williams and Kaufmann (2012) specifically investigated the reliability of the GNAT. They recommend a minimum of 40 trials per block for minimally acceptable reliability and at least 80 trials per block for good reliability. They argue that the GNAT is a valuable tool with many advantages and should be used in further research. (Williams & Kaufmann, 2012). By omitting a comparison concept (unlike the IAT), the GNAT can use distractor items more flexibly, allowing for a direct assessment of the attitude (Nosek & Banaji, 2001). “*In addition, the GNAT may be less susceptible to errors introduced by term valence and less biased by response criteria than reaction time-based techniques.*” (Boldero et al., 2007, p. 354).

The convergence between implicit and explicit personality traits was examined by Boldero et al. (2007). They support the reliability and convergent validity of the GNAT when controlling the systematic variance of the GNAT. However, as their explicit measure was conducted before the implicit GNAT, it is possible that this inflated their associations and, thus, the correlation between both measures (Boldero et al., 2007).

Unfortunately, sufficient research in the moral domain, including the GNAT, has not yet been conducted. Previous studies focused on predictions about moral behavior with implicit measurement methods, like the IAT (Perugini & Leone, 2009). Another study tried to measure the moral identity with the IAT (Johnston et al., 2013). The only known publication of an application of the GNAT to measure the implicit moral self-image is a dissertation by Ferguson (2018), who failed to show moral balancing effects.

Implicit measures, such as the IAT and GNAT, can easily be implemented into software and be used as a portable version on mobile devices, thus overcoming the limitation of requiring a local computer (Bassett & Dabbs, 2005; Dabbs et al., 2003). This provides several advantages. Firstly, it tests participants in a more natural setting outside a laboratory (Bassett & Dabbs, 2005; Dabbs et al., 2003). This real-life setting could lower the feeling of being observed during the experiment and lead to more honest and impulsive answers. Secondly, Bassett and Dabbs (2005) argue that portable versions could be used to measure malleable attitudes at different times and important events. Thirdly, having a mobile and portable version of these tests could help to reach populations that would usually not participate in laboratory experiments. Fourthly, more people could participate because the effort needed is much less, as they are no longer required to go to the laboratory (Bassett & Dabbs, 2005).

3. Methodology

3.1. Research Design

This bachelor's thesis aims to apply an existing implicit measurement method, the Go/No-Go Association Task (GNAT), to measure the participants' implicit moral self-image and analyze these results for a correlation with their explicit moral self-image. It intends to investigate the effectiveness of the GNAT with the convergent validity between the implicit and explicit measures. As explained in Chapter 2, most of the previous research to measure the moral self-image is not based on the GNAT (e.g., Johnston et al., 2013; Perugini and Leone, 2009).

The underlying parameters of the performed experiment and the test setup were as follows:

3.1.1. Experimental Design

Objective & Material & Groups & Variables

Objective: Measure the implicit moral self-image and analyze the correlation with the explicit moral self-image (convergent validity).

Material and Groups: Inspired by Ferguson (2018), who failed to provide significant evidence for moral balancing using the GNAT, her experiment was reproduced in group A using the exact same stimuli (six words for GOOD/BAD and four words for ME/OTHER) to be consistent with her research method. In group B, the list of words was extended for potentially stronger effects.

Group A: These stimuli words were replicated from Ferguson (2018):

ME: *me, I, my, myself*

OTHER: *other, others, them, they*

GOOD: *good, honest, faithful, modest, sincere, altruist*

BAD: *bad, dishonest, deceptive, pretentious, arrogant, cheater*

Group B: The extended stimuli words (attributes) were selected from different scientific papers and measured in a pre-study according to their evaluative intensity (see Appendix VI Pre-Study GOOD & BAD). This group used 20 stimuli words for each attribute and 15 for each concept category (ME & OTHER). To be consistent with Ferguson (2018) and the previous group A, the 40 attribute stimuli (57.14%) and 30 concept category stimuli (42.86%) in relation to the total stimuli of 70 in group B was almost equal to the 12 attribute stimuli (60%) and 8 concept category stimuli (40%) initially introduced by Ferguson (2018).

The used stimuli words in group B were:

ME: *me, I, my, myself, mine, self, personally, oneself, person, intrinsic, own, individual, ego, inner essence, inner self*

Table 1: Possible outcomes of the Go/No-Go Association Task, inspired by “Trial of yes-no experiment” in Macmillan (2002)

		Response			
		Go		No-Go	
Stimulus	Signal (target item)	Hit	○	Miss	×
	Noise (distractor item)	False Alarm	×	Correct Rejection	○

OTHER: *other, others, them, they, their, themselves, theirs, his, him, her, anybody, anyone, those people, persons, the individuals*

GOOD: *caring, fair, compassionate, friendly, hard-working, generous, helpful, kind, honest, faithful, altruist, modest, sincere, genuine, joyful, patient, grateful, loyal, forgiving, respectful*

BAD: *hostile, unfair, lazy, unhelpful, ruthless, selfish, evil, brutal, hateful, angry, impatient, bad, dishonest, deceptive, pretentious, arrogant, cheater, disrespectful, disloyal, egocentric*

Variables: The stimuli word lists (group A or B) served as independent variables. The dependent variables were Hit/False-Alarm rates.

Blocks & Trials & Stimuli

The GNAT was divided into two blocks. The target category ME was paired with the attribute GOOD in the starting block. In the second block, the same target category, ME, was paired with the opposite attribute, BAD.

Each block comprised a total of 96 trials for group A or 86 trials for group B. Both blocks started with 16 practice trials (not considered in the analysis), followed by a reminder screen, before proceeding to the 80 critical trials (considered in the analysis) for group A or 70 critical trials for group B. A trial started when a stimulus word from one of the categories (ME, OTHER, GOOD, BAD) emerged on the screen. It ended when the word disappeared. As a constant reminder of the current combination (pairing) in each block, labels for the target category (ME) and the attribute (GOOD or BAD) remained on the screen's upper left and right corners. The labels and stimuli items were displayed in black font against a white screen.

The participants were advised to either (1) give a Go response by quickly pressing the space bar if the stimulus word displayed could be categorized into one of the two labeled categories (signal item) or (2) refrain from pressing any key (No-Go response) for words that could not be categorized (noise items). The stimulus word appeared in the center of the screen and remained visible until the response deadline was reached or a key was pressed. The subsequent trial started when the participant pressed the space bar or after the response time ran out. Like Nosek and Banaji (2001), the opposing category (OTHER) or the alternate attribute served as distracter trials (noise). For example, when GOOD was the

signal, BAD was the noise, and vice versa. A signal-to-noise ratio of 1:1 was held constant for all trials and both groups.

The 20 stimuli words Ferguson (2018) used for the critical trials in group A were selected randomly. Each word was repeated four times for a total of 80 trials, which was in the range of 50 to 80, yielding sufficient and good reliability, as Williams and Kaufmann (2012) recommended. For group B, the stimuli items were chosen randomly and selected without repetition from the four categories (ME, OTHER, GOOD, BAD) to reach 70 trials. Within both groups, each block (pairing) consisted of an equal number of words in order to minimize learning effects, as was again advised by Williams and Kaufmann (2012). The critical trials used the complete set of stimuli words, which were selected randomly and appeared in random order. Additionally, for the word list in group B, a small-scale pre-study (Appendix VI Pre-Study GOOD & BAD) on the evaluative intensity of these words was conducted beforehand. This ensured that the words' evaluative intensity was a) strong enough to yield sufficient results and b) similar between the words, which was strongly suggested by Nosek and Banaji (2001).

Response Deadline & Feedback

The participants had to categorize the stimuli words as quickly and accurately as possible during the short time displayed on the screen. The response deadline was constant at 700 milliseconds (ms) across all trials and blocks, following the recommended range of 500ms to 850ms by Nosek and Banaji (2001). The interstimulus interval between two trials was held constant at 500ms.

During this interstimulus interval, immediate feedback on performance accuracy was provided. Trials where signal items were accurately identified (Hit) or noise items were correctly ignored (Correct Rejection) were recorded as correct responses, indicated by a green “O” appearing. Trials were marked as errors when noise items were mistakenly identified as signals (False Alarm) or signal items were overlooked (Miss). For these trials, a red “X” was displayed in the center of the screen after the stimulus item disappeared. The possible outcomes and corresponding feedback are visualized in Table 1.

3.1.2. Statistical Analysis

The statistical analysis of the GNAT was based on the signal detection theory, first introduced by Green and Swets (1966), as cited in Macmillan (2002), and as previous experiments using the GNAT already have done (Ferguson, 2018;

Nosek & Banaji, 2001; Teachman, 2007).

An essential part of the signal detection theory is the calculation of the sensitivity, described as d' . It indicates the participant's ability to differentiate signal (target items) from noise (distractor items) (Green & Swets, 1966; Macmillan, 2002; Nosek & Banaji, 2001). D -prime (d'), based on the signal detection theory and applied in the GNAT by Nosek and Banaji (2001), is calculated for each block (ME + GOOD and ME + BAD) with the following formula:

$$d' = Z\left(\frac{\text{Hits}}{\text{Signal items}}\right) - Z\left(\frac{\text{False Alarms}}{\text{Noise items}}\right)$$

It calculates the difference between the standardized Z-score of the ratio of correct Go responses (Hits) to all signal items and the standardized Z-score of the ratio of Go responses for noise items (False Alarms) to all noise items (Macmillan, 2002). D -prime (d') only considers the participant's Go responses, as only the absolute number of Hits and the absolute number of False Alarms relative to the total amount of signal or noise items in the experiment are included in the formula. Generally, a higher d' reflects a higher sensitivity. This means that the participant could discriminate signal items from noise items more easily. In other words, he had more correct Go responses (Hits) than Go responses for distractor items (False Alarms).

Implicit Moral Self-Image Score

The implicit moral self-image score or d' (GNAT) determined by the GNAT was defined as the difference in associative strength (sensitivity) between the two blocks (Nosek & Banaji, 2001). The d' of the second pairing (ME + BAD) was subtracted from the d' of the first pairing (ME + GOOD), which resulted in the final implicit moral self-image score d' (GNAT)⁴:

$$\text{implicit moral self - image score :} \\ d'(\text{GNAT}) = d'(\text{ME} + \text{GOOD}) - d'(\text{ME} + \text{BAD})$$

A positive implicit moral self-image score indicates a positive moral self-image since the associations between ME + GOOD are stronger than between ME + BAD. In contrast, a negative implicit moral self-image score is interpreted as a negative moral self-image because of a stronger association between ME + BAD than between ME + GOOD (Ferguson, 2018).

Explicit Moral Self-Image Score

The explicit moral self-image score was determined with the questionnaire (see Appendix VII Explicit Moral Self-Image Questionnaire) introduced by Jordan et al. (2015) to

investigate the relationship and correlation between explicit and implicit moral self-image. Participants were asked about nine moral traits on a nine-point Likert scale, where 1: [I am] much less "moral trait" (e.g., caring) than the person I want to be and 9: [I am] much more "moral trait" (e.g., caring) than the person I want to be. The explicit moral self-image score was calculated by taking the average of all responses in the explicit moral self-image questionnaire. This score ranged from 1 to 9, with 5 demonstrating a neutral explicit moral self-image, lower numbers indicating a lower explicit moral self-image, and higher numbers indicating a higher explicit moral self-image, respectively.

Pearson Correlation Coefficient

The (Pearson) correlation coefficient (r) between each group's implicit moral self-image scores, based on response rates, and the corresponding explicit moral self-image scores, was determined to evaluate the convergent validity between both measurement methods. This coefficient is standardized from -1, indicating a perfect negative linear relationship, to 1, indicating a perfect positive linear relationship, and 0, indicating no linear correlation (Fahrmeir et al., 2016). For the formula, see VIII Formula of the Pearson Correlation Coefficient.

3.2. Participants

124 out of 206 test-takers successfully finished the experiment, corresponding to a conversion rate of 60.19%. Out of these 124 participants, 56 were assigned to group A and 68 to group B at random. This sample of 124 participants consisted of 71 females (57.26%), 52 males (41.93%), and 1 person not indicating their gender (0.81%). The participants' average age was 30.08 years. 84 people (67.74%) participated on mobile devices, whereas 40 (32.26%) participated on laptops or desktops.

3.3. Procedure

The participants received a link to the oTree website, an open-source experiment platform (Chen et al., 2016). They could access the experiment through the internet using their laptop as a stationary device or mobile devices such as smartphones or tablets. Using a cookie on the oTree website ensured that participants were only allowed to participate once per device. After opening this link, the participants were forwarded to the welcome page, where they received the first information about this experiment. They were told that it was part of a bachelor's thesis and that their contribution would help assess error rates and reaction times when categorizing various English words. On the next page, they received detailed instructions about the test procedure.

For the GNAT, the participants were randomly assigned to one of the two groups and, thus, different word lists. After that, they saw an overview of the current labels in the first block (ME + GOOD) and the list of words. They started immediately with a training test, the first 16 practice trials.

⁴ For further clarification: d' (GNAT), also known as implicit moral self-image score based on response rates, represents the final score of the GNAT, which was calculated as the difference between the d' of the different blocks (pairings) (see formula).

Table 2: GNAT procedure in the experiment

<ul style="list-style-type: none"> • Welcome • Instructions: How It Works 	
Group A	Group B
Block 1: (ME + GOOD) <ul style="list-style-type: none"> • Overview (labels and word list) • 16 Practice Trials • Reminder • 80 Critical Trials 	Block 1: (ME + GOOD) <ul style="list-style-type: none"> • Overview (labels and word list) • 16 Practice Trials • Reminder • 70 Critical Trials
Block 2: (ME + BAD) <ul style="list-style-type: none"> • Overview (labels and word list) • 16 Practice Trials • Reminder • 80 Critical Trials 	Block 1: (ME + BAD) <ul style="list-style-type: none"> • Overview (labels and word list) • 16 Practice Trials • Reminder • 70 Critical Trials
<ul style="list-style-type: none"> • Explicit Moral Self-Image Questionnaire • Demographic Questions 	

After finishing a task, the participants saw the number of errors they had made. They were directed to a page reminding them of the current pairing and word list, that they should answer as quickly and accurately as possible, and that the next part (critical trials) counts toward their scores. For the second block with the new pairing ME + BAD, the same procedure started again, consisting of the first 16 practice trials, the reminder, and the critical trials.

After finishing the second block, participants were explicitly asked about their moral self-image using the scale developed by Jordan et al. (2015). They were asked about nine moral traits on a nine-point Likert scale, where 1: [I am] much less “moral trait” (e.g., caring) than the person I want to be and 9: [I am] much more “moral trait” (e.g., caring) than the person I want to be. Finally, some demographic questions, including gender, age, educational level, employment status, first language, and English proficiency, as well as the occurrence of technical problems, were asked. A graphical representation of the procedure for both groups is visualized in Table 2.

3.4. Hypotheses

The following hypotheses were derived to measure the effectiveness of the GNAT according to its convergent validity with an explicit moral self-image questionnaire:

Hypothesis 1: The participants’ implicit moral self-image scores (based on response rates) will **correlate positively** ($r > 0$) with the explicit moral self-image scores (questionnaire-based) in both groups.

This hypothesis assumes that both implicit and explicit moral self-image reflect the same construct but are measured through different methods, following Johnston et al. (2013) and Fazio and Olson (2003). Therefore, a positive correlation between both measures is expected, as the scores may vary

but do not need to be congruent. A person with a positive implicit moral self-image in the GNAT should also indicate a positive explicit moral self-image in the questionnaire, and vice versa.

Hypothesis 2: The **correlation** between the participants’ implicit moral self-image scores and explicit moral self-image scores will be **higher in group B than in group A**.

This hypothesis assumes that the word list without repetition of group B would be more effective (and yield higher correlation) in the GNAT than the smaller, repeated word list of group A. It is expected that the repetition of words will lead to participants’ better memorization during the first block, allowing for improvements and better performance in the second block, reducing the implicit moral self-image score. This would reduce the correlation coefficient compared to group B, where words are not repeated.

Table 3: Exclusion criteria for the standard sample

Exclusion Criteria
1. No recorded data
2. No response was given during the trials
3. Reaction times < 250ms
4. Reaction times > 735ms
5. $d'(ME + GOOD) \leq 0$ or $d'(ME + BAD) \leq 0$
6. Basic English proficiency
7. Indication of technical issues
8. Perfect Hit rate
9. No False Alarms

4. Results

4.1. Exclusion Criteria

Some participants had to be excluded from the analysis to ensure a valid and valuable data set (standard sample). Therefore, the data set was manually revised, especially due to technical difficulties during data collection, such as the failure to record responses and/or reaction times as well as heavy deviations from the response deadlines (e.g., speeding up the appearing words (up to 1ms) or very long loading times (up to 1,080ms)).

The following exclusion criteria for the standard sample were defined (and visualized in Table 3):

All participants whose responses were not recorded were excluded. Either the server failed to measure the reaction times for some trials, or the participants failed to respond correctly by not giving Go responses (in one block or throughout the experiment).

Furthermore, all participants with response times below 250 milliseconds (< 250ms) and above 735 milliseconds (> 735ms), to still account for some processing time of the server, were dropped. Greenwald et al. (2003) suggest excluding reaction times below 300 milliseconds (< 300ms), which could indicate random responding, but due to the advanced digitalization, frequent use of electronic devices nowadays, and the relatively young sample (average age of 30.08 years), response latencies within the range of 250ms to 735ms were included.

Additionally, following the argument by Nosek and Banaji (2001), all d' -primes (d') (based on response rates) below 0 were excluded as they indicate that the participant was not following the instructions properly or was unable to distinguish any signal from noise. A value of 0 equals chance (Nosek & Banaji, 2001).

All participants had to have at least intermediate English proficiency to be considered. Therefore, all participants who indicated basic English skills were dropped. They were also excluded if they indicated technical issues in the final demographic questionnaire.

Further, due to problems with the mathematical Z-score calculation, all participants with perfect Hit rates (a participant correctly responded to all signal items) or with no False Alarms (a participant never gave a Go response for distractor items) were excluded from the primary analysis. Nevertheless, the effects of a potential correction when calculating d' were investigated.

4.2. Participants per Group

Group A (word list by Ferguson (2018)):

After the strict exclusion, 30 participants of the initial 56 people finishing the experiment in group A were considered (conversion rate of 53.57%). The sample consisted of 17 females (56.67%) and 13 males (43.33%), with an overall average age of 29.43 years.

Most people were students (56.67%). The rest was either full-time (30%) or part-time employed (13.33%). As their highest completed level of education, nine people indicated high school degrees (30%), ten had a Bachelor's degree (33.33%), and 36.67% had a Master's degree. The majority spoke German as their first language (76.67%), two spoke Portuguese (6.67%), and one person spoke English, Dutch, Spanish, and Finnish, respectively. One person stated being bilingual in German and French (3.33%). 73.33% of participants experimented on a mobile device, while 26.67% used stationary devices.

Group B (extended word list):

Following the same exclusion criteria, 38 participants of the initial 68 people finishing group B were considered (conversion rate of 55.88%). The sample consists of 21 females (55.26%) and 17 males (44.74%) with an overall average age of 30.82 years.

Half of the participants were students (50%). The rest were either full-time (36.84%), part-time (5.26%) or self-employed (5.26%), and one person had already retired (2.63%). As their highest completed level of education, eight people indicated high school degrees (21.05%), twenty people had a Bachelor's degree (52.63%), and 26.32% had a Master's degree. Most people spoke German as their first language (76.32%), five people were native English speakers (13.16%), two people spoke Spanish (5.26%), and one person spoke Ukrainian (2.63%). One person stated to be bilingual in German and English. 69.23% of participants performed the experiment on mobile devices, and 28.25% on desktops or laptops.

4.3. Implicit Moral Self-Image Score Based on Response Rates

Signal detection theory proposes calculating participant sensitivity (d') for each block. It subtracts the standardized Z-scores of the ratio of Go responses for noise items (False Alarms) to all noise items from the standardized Z-score of the ratio of correct Go responses (Hits) to all signal items (Macmillan, 2002). Applying this theory to the GNAT, it is assumed that participants are more sensitive and can discriminate signal items from noise items more easily when the association of the two signal components is more positive than a negative or neutral association (Nosek & Banaji, 2001). "Greater sensitivity indicates a stronger association between the target category and attribute. This association is defined to be a measure of automatic attitude toward the target category." (Nosek & Banaji, 2001, p. 635).

To calculate the implicit moral self-image score with the GNAT, the approach of Ferguson (2018) and Nosek and Banaji (2001) was followed by comparing the d' of both pairings for each participant:

Implicit moral self – image score :

$$d'(\text{GNAT}) = d'(\text{ME} + \text{GOOD}) - d'(\text{ME} + \text{BAD})$$

According to Ferguson (2018), a positive implicit moral self-image score indicates a positive moral self-image (the associations between ME + GOOD are stronger than between ME + BAD), whereas a negative implicit moral self-image score is interpreted as a negative moral self-image (stronger association between ME + BAD than between ME + GOOD).

4.3.1. Results of Standard Sample

Table 4 shows all characteristics of the obtained implicit moral self-image scores for both groups and compares the Pearson correlation coefficients of both groups.

Table 4: Characteristics of implicit moral self-image scores, based on response rates, and Pearson correlation coefficients for group A and group B with strict exclusion criteria

	Group A	Group B
Max	1.3804	1.5475
Min	-1.4816	-0.6629
Mean	0.2360	0.4154
Median	0.2783	0.3588
Pearson correlation coefficient	-0.0078	0.4870

Group A

In group A, the final implicit moral self-image scores, calculated as the participants' difference in association strength (d') of both blocks, ranged from a score of -1.4816 to 1.3804, with a mean of 0.2360 and a median of 0.2783. Eleven participants (36.67%) performed better in the ME + BAD block, indicating a negative implicit moral self-image. The questionnaire-based average explicit moral self-image score was obtained as 5.2852 in group A.

Correlating the implicit moral self-image scores with the explicit moral self-image scores conveyed through the questionnaire after the GNAT yielded a Pearson correlation coefficient of $r = -0.0078$. This coefficient does not indicate a linear correlation between the explicit and implicit moral self-image scores. Counterintuitively, it may suggest a very slight negative tendency. This means that people who performed better in the ME + GOOD pairing of the GNAT, relative to the ME + BAD pairing, provided answers in the questionnaire indicating a lower explicit moral self-image score and vice versa. This does not support hypothesis 1 of the correlation coefficient being positive for all groups ($r > 0$). Figure 1 depicts the distribution of values and ultimately demonstrates the correlation (dotted line).

Group B

The implicit moral self-image scores conducted through the GNAT for group B had a mean of 0.4154, a median of

0.4154, and ranged from -0.6629 to 1.5475. Only nine participants (23.68%) had a negative implicit moral self-image score and thus a higher association of ME + BAD. The average explicit moral self-image score in this group was 5.2281.

The Pearson correlation coefficient between implicit moral self-image scores and explicit moral self-image scores was calculated as $r = 0.4870$, which indicates a moderate positive linear relationship. This finding suggests that people who performed better in ME + GOOD pairing had a higher average score for the explicit moral self-image. This supports the first hypothesis of a positive correlation ($r > 0$) and the effectiveness of the GNAT. Figure 2 depicts the distribution of values and ultimately demonstrates the correlation (dotted line) for group B. A similar pattern but different ranges on the x-axis can be observed in Figure 1 and Figure 2. When comparing both correlation coefficients, it becomes evident that group B correlates higher than group A, which supports hypothesis 2.

4.3.2. Correction of 0.5 for all Perfect Responses

As previously stated, all participants with perfect Hit rates and no False Alarms (perfect responses) were excluded from the data set due to problems when calculating the Z-scores and d' of each block. Nonetheless, the effects of a correction suggested by Kadlec (1999) and Macmillan (2002) on the correlation coefficient for both groups were investigated. This approach involves correcting the absolute numbers of Hits or False Alarms by either adding 0.5 to no False Alarms or subtracting 0.5 from perfect Hit rates, which yielded 39.5 Hits in group A and 34.5 Hits for group B. Correcting these perfect response rates allowed four more participants to be included in group A and two more in group B.

This resulted in Pearson correlation coefficients of $r = 0.0106$ for group A and $r = 0.4384$ for group B. In this case, both groups demonstrate a positive correlation supporting hypothesis 1. The coefficient increases for group A and decreases for group B compared to the standard sample, as shown in Table 5. However, in support of hypothesis 2, the correlation coefficient of group B is still higher than group A's.

Table 5: Comparison of different Pearson correlation coefficients before and after sample correction for groups A and B

Pearson correlation coefficient	Group A	Group B
Standard sample (before correction)	-0.0078	0.4870
New sample, including participants with perfect responses (after correction)	0.0106	0.4384



Figure 1: Correlation (dotted line) between the implicit moral self-image scores, based on response rates, and the explicit moral self-image scores for group A

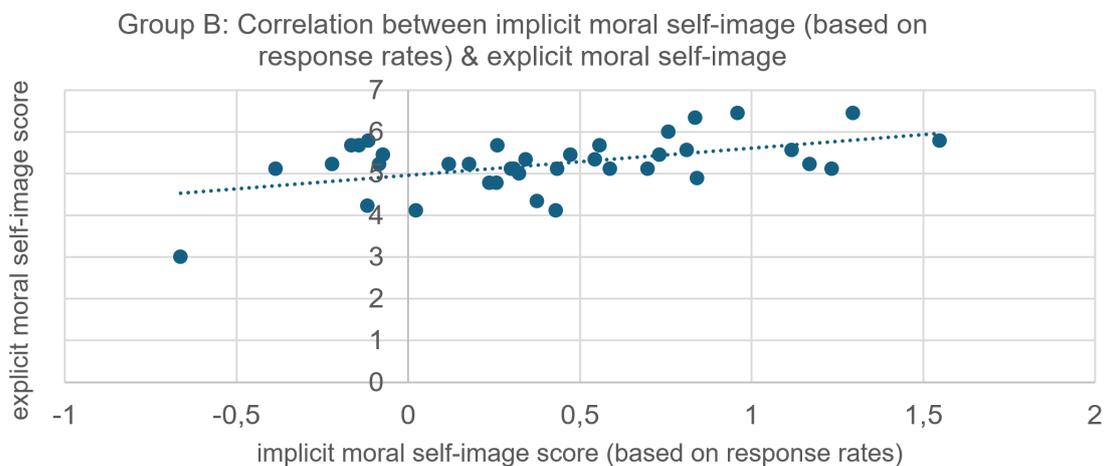


Figure 2: Correlation (dotted line) between the implicit moral self-image scores, based on response rates, and the explicit moral self-image scores for group B

5. Discussion

5.1. Key Findings

This study was designed to investigate the effectiveness of the GNAT for measuring the moral self-image with its convergent validity by relating it to the well-established explicit moral self-image questionnaire by Jordan et al. (2015). The indicator considered in this analysis is the Pearson correlation coefficient (r), where a high correlation indicates a high convergent validity. The implicit moral self-image was analyzed with a focus on the response rates using the signal detection theory by Green and Swets (1966), as cited in Macmillan (2002).

This bachelor's thesis is the second known attempt to measure the implicit moral self-image with the GNAT. Given that the first attempt by Ferguson (2018) was unsuccessful in showing significant results for moral balancing and was somewhat opaque and vague about the detailed research design, this was replicated in group A but extended in group B. Based on literature, Ferguson's (2018) list of words was extended with more synonyms for group B (see Appendix VI

Pre-Study GOOD & BAD) to further explore the effect of different stimuli words in the GNAT. Another crucial difference between both groups was the repetition of stimuli. Group A only used 20 stimuli words (10 signals and 10 noises), which were each repeated four times to yield a total of 80 critical trials, whereas group B used 70 stimuli words (35 signals and 35 noises) for 70 trials without repetition.

In hypothesis 1, both groups were expected to have a positive correlation between the implicit and explicit moral self-image scores ($r > 0$). This assessment was based on the idea that implicit and explicit moral self-image explain the same construct and are only obtained by different measurement methods, following Johnston et al. (2013) and Fazio and Olson (2003). Consequently, this would lead to a positive correlation, even though the scores allow for some deviation and do not necessarily need to be completely congruent.

This hypothesis 1 has been proven wrong for group A, as the first group with fewer stimuli had a negative correlation coefficient of $r = -0.0078$ in the standard sample. This negative relationship could be explained, as it seems reasonable

that participants who performed better in the second block (ME + BAD) had already seen all stimuli at least four times before and memorized the stimuli used in the first block. This prior “training” in the first block might have caused a higher d' in the second block (ME + BAD), which led to a negative implicit moral self-image score in the GNAT. With this improved personal performance in the second block (ME + BAD), participants could have bolstered their moral self-image and could have reported a higher score in the questionnaire, which led to a negative correlation. This appears plausible since they received immediate feedback and saw the total number of errors after each block. The same mechanism could be applied to participants who performed even worse in the second block. Participants not fulfilling their target in the second block, e.g., not outperforming their error rate from the first block, might be disappointed and state a negative explicit moral self-image. As the negative tendency was extremely small, this influence of personal performance on the explicit questionnaire was insignificant and could not be observed after the correction. However, with a convergent validity of $r = -0.0078$ for group A or even after the sample correction of $r = 0.0106$, both coefficients were very close to zero, indicating no significant correlation and, thus, no convergent validity. The absence of (sufficient) convergent validity might imply that both measures do not assess the same construct or that either of the two measures is flawed (Cunningham et al., 2001).

In contrast, group B had a Pearson coefficient of $r = 0.4870$ (even after correction $r = 0.4384$), which makes a particular relationship and a moderate convergent validity between both measures evident for this group. This further proves that the GNAT, as an implicit measure, is not flawed and holds good convergent validity, which can be interpreted as proof that implicit and explicit measures capture the same construct (to a certain extent). This promising relationship could be understood as even more substantial and already as significant since it is much higher than the previous standard for implicit measures (usually around 0.2 to 0.3 (Perugini, 2005)). Cunningham et al. (2001) noted that the extent of measurement error in any measure establishes a natural upper limit for correlations. They interpreted their correlation of $r = 0.35$ between the implicit IAT and an explicit scale already as significant and argued that “*the two sets of measures are correlated, yet distinct.*” (Cunningham et al., 2001, p. 167).

For hypothesis 2, it was expected that the correlation between the implicit moral self-image scores and the explicit moral self-image scores (based on the questionnaire) for group B would be higher than that of group A. This hypothesis has been proven right due to a significantly higher correlation coefficient of $r = 0.4870$ for group B compared to $r = -0.0078$ for group A. Even when considering the correction for all perfect responses, the same effect could be observed with a correlation coefficient of $r = 0.4384$ and $r = 0.0106$ for group B and group A, respectively.

This could also be explained by stronger learning effects in group A, as the same 20 stimuli were repeated four times

in each block. This was not the case for group B. Each stimulus was only shown once within a block due to the extended word list. Additionally, in relative terms, more participants with a negative implicit moral self-image score were observed for group A (36.67%) than for group B (23.68%). This could be because participants in group A had fewer stimuli and more time to learn, which gave them an advantage for the second block, where they performed better in 36.67% of all cases. This percentage is lower in group B, at 23.68%. Therefore, it can be assumed that learning effects in group B were not that strong, which led to a more significant correlation between implicit and explicit measures. In addition, learning effects due to unequal block length were avoided in the experimental design, as both blocks in each group had the same length; hence, this should not cause this significant difference in correlation between both groups (Williams & Kaufmann, 2012).

An alternative explanation for the higher correlation could be the higher percentage of native English speakers in group B (13.16%) compared to group A (3.33%). Indeed, a native English speaker would react faster and more accurately than a native German speaker. However, assuming that the characteristics of each participant, for instance, language or age, were the same in both blocks, this should not affect the overall results. This is because the response rate of each participant was calculated into an individual d' for each block by only comparing the response rates within each participant (identical person). The d' of a native English speaker could then be generally higher in each block compared to the d' of a German speaker, but after calculating the difference between both blocks (the implicit moral self-image score), this effect can be neglected. Therefore, language, age, and educational level should not affect the overall results in both groups.

As mentioned, previous literature usually obtained lower correlation coefficients (generally between 0.2 and 0.3) and explained this as the natural difference between explicit (relying on self-reporting) and implicit (relying on the associative strength and measured with reaction times or error rates) measurement methods (Cunningham et al., 2001; Perugini, 2005). Assuming the correlation is a perfect indicator of the underlying system, one could argue that a lower correlation would prove a double dissociation system, where explicit and implicit attitudes exist and describe different evaluations. This only applies if the assumption holds that explicit attitudes are only measured by explicit measures and implicit attitudes only by implicit measures. In contrast, a higher correlation would support the existence of an additive pattern, implying that explicit and implicit measures describe the same attitude, as both measures yield similar results. As there is no clear consensus among researchers, it is difficult to come to a conclusion. However, one could believe that “*implicit and explicit attitudes can be best understood as implicit or explicit measures of the same attitude.*” (Perugini, 2005, p. 31), given the promising correlation in group B.

The initial research question: “To what extent can the Go/No-Go Association Task (GNAT) be effectively applied

to measure moral self-image, and is there a correlation between the outcomes of this method and the explicit moral self-image?" should be answered separately for both groups. Group A has shown little, even negative, to no correlation between the two measures, which indicates an ineffective application of the GNAT in measuring the moral self-image. For group B, there is, in fact, a significant positive correlation between the implicit and explicit measures, indicating an effective application because learning effects were avoided by using various stimuli without repetitions.

This thesis has closed the gap in previous research, as it has proven that the GNAT is an effective instrument for measuring the moral self-image if learning effects are considered in the research design. It suggests utilizing various stimuli words without repetition instead of repeating fewer stimuli words. This finding has highlighted the extraordinary importance of the experimental design for yielding significant results. However, the GNAT still has some limitations, and further research investigating the GNAT's ability to predict behavior and convergent validity is highly recommended, whereas the first results were already quite promising. This thesis was the first study conducted to assess the convergent validity of the GNAT for capturing the moral self-image.

5.2. Limitation of Results & Suggestions for Future Research

It is important to emphasize the limitations of this research, especially to enable future research. First of all, investigating the correlation between explicit and implicit measures could be misleading and ambiguous due to various explanations for both attitudes and missing consensus about its significance. According to Perugini (2005), a low correlation could suggest insufficient convergence validity between the two measurement methods. However, it could also be seen as evidence for the double dissociation, supporting the theory that a dual system of attitudes exists and that those attitudes are unrelated. Instead, it might be more expressive to separately investigate the capability of both measures to predict behavior (Perugini, 2005) and explore this in further experiments. Validating the GNAT as a moral self-image measure according to its ability to predict behavior would provide further insights and could support its usefulness (Perugini, 2005).

Nevertheless, Carlson and Herdman (2012) generally advised using measures with convergent validities above 0.7 ($r > 0.7$) and avoiding those below 0.5 ($r < 0.5$). In the context of implicit measures and especially regarding the moral self-image, one might have to deviate from this suggestion, as there are various reasons for the low correlation. More research should be conducted to review these recommendations.

Future research could extend the experiment over a longer timeframe and ask participants about their moral self-image before and after specific moral or immoral actions. Since the measured moral self-image is a snapshot (in daily life) and can deviate in time (Jordan et al., 2015; Monin & Jordan, 2009), tests could be conducted at regular

intervals with the identical test group to analyze the ranges of variation around their fixed personal reference, potentially initiating moral balancing behavior (Jordan et al., 2015; Nisan, 1990).

Regarding the experimental design, further investigation is needed about the effect of the block order on the moral self-image, as previous research has already suggested that it might have an impact (Greenwald et al., 1998, 2003; Nosek & Banaji, 2001). It would be interesting to see if a reversed order of the blocks yields the same results and convergent validity. An extensive experiment should investigate the occurrence of learning effects due to increased familiarity with the procedure and stimuli words in the second block. This could be done by changing the order of the blocks ME + GOOD and ME + BAD in different groups.

When investigating the convergent validity, an additional aspect is changing the order of the explicit questionnaire. Asking participants explicitly about their moral self-image before and after the GNAT could examine potential priming effects (of both explicit questionnaire and implicit GNAT), eventually inflating the correlation (Boldero et al., 2007). Further, including the questionnaire twice (before and after the GNAT) could provide insights into how the individual's performance in the GNAT changes the explicit moral self-image.

Additionally, it would make sense to expand this sample and conduct the experiment with more participants. Simple deviations already greatly impacted the coefficients obtained in the current sample of 30 (group A) to 38 participants (group B). Additionally, the uneven size of both groups raised concerns about the comparison. Most participants were native German speakers (76.67% in group A and 76.32% in group B), which differentiated from previous experiments, where only native English speakers were asked to perform the GNAT in English (Ferguson, 2018; Nosek & Banaji, 2001; Williams & Kaufmann, 2012). Repeating the same experiment in the first language of each participant could eventually provide more robust results, as the implicit association could be much stronger, investigating more "natural" and more accessible associations.

In addition, the portable nature of the GNAT (Bassett & Dabbs, 2005), as conducted in this online experiment, allowed for measuring the moral self-image of participants in everyday life. This setting enabled participants to perform the experiment in a familiar environment, without mandatory seminars, away from an unpleasant and stressful (graded) laboratory atmosphere, and whenever they felt ready and comfortable to process the test. Participants were only instructed to perform the test to support a bachelor's thesis in assessing error rates and reaction times when categorizing English words without telling them the true purpose of this experiment. This more relaxing environment might have supported less biased, unconscious responses and yielded real-life insights into participants' moral self-image without them being aware of it. However, this could also have led to participants not taking the test seriously and giving less concentrated responses than in a controlled

laboratory environment, which made it harder to compare the results with previous studies.

Furthermore, the differences in results obtained through mobile and stationary devices were not investigated due to the limited scope of this bachelor's thesis and the small sample size. Examining this further with a larger sample is highly recommended, as the experimental Go response differed for both versions. Participants using touch screens had to tap exactly on the word, whereas participants using laptops had to press the space bar.

Moreover, due to the limited scope of this bachelor's thesis and the partially experienced technical discrepancies, a detailed analysis of the reaction times (response latencies), according to Cohen (2013), was omitted. An analysis applying the signal detection theory based on reaction times is proposed in Appendix V.IV Implicit Moral Self-Image Score Based on Reaction Times. However, an analysis based on Cohen (2013) and the often-used algorithm developed by Greenwald et al. (2003), relying on mean latencies and standard deviation of all reaction times, is strongly recommended in the future as it could hold important insights and support the GNAT's validity by confirming the results and correlations obtained.

5.3. Future Implications

After introducing the IAT and the GNAT, a new field of research opened up for assessing people's implicit attitudes. This could not only be used for psychological counseling (Asendorpf et al., 2002) but can also provide valuable and hidden insights into behavioral economics and for different economic stakeholders, e.g., companies, employees, investors, and policymakers.

Companies could use implicit measures to gain deeper insights into their stakeholders and to predict consumer behavior, especially when bringing this into the context of the moral balancing theory. They could exploit the knowledge about the compensatory balancing behavior of people after an initial immoral act. For example, companies could offer customers the option to donate to a prosocial charity after a self-centered purchase (e.g., a carbon-intensive flight for vacation), as adapted from Schlegelmilch and Simbrunner (2019). This allows customers to compensate for a decreased moral account with a subsequent good action (moral cleansing). Moral licensing effects could be utilized, for instance, by allowing people to demonstrate moral behavior (e.g., donating to a charity) before offering them a carbon-intensive flight (adapted from Schlegelmilch and Simbrunner (2019)). Both effects influence the customer's purchase decision, giving them a better feeling after balancing or justifying their self-centered purchase, potentially leading to a higher demand in the future. Additionally, "[i]mplementing donation options in a web shop is an easy way for a company to signal that it is socially responsible." (Schlegelmilch & Simbrunner, 2019, p. 551). In Marketing, companies could cluster their target group, as people with higher importance on their morality have a higher incentive to maintain their moral self-image and are more likely to be influenced by (negative)

deviations (Aquino & Reed, 2002; Monin & Jordan, 2009). If a person with a high importance on morality performs an immoral act, it has a larger negative impact on his moral self-image, and he is more likely to compensate for it afterward (Monin & Jordan, 2009). This could be utilized for personalized advertisements, pricing, and influencing people. As the GNAT reveals deeper, unconscious attitudes and feelings toward brands or products, it could be applied to more generic and strategic decisions, such as product development process, brand positioning, pricing, and improved market research.

Assessing the implicit moral self-image may serve as an efficient tool in the recruitment process (Asendorpf et al., 2002), potentially revealing the implicit biases and attitudes of both recruiters and prospective employees. This could be a helpful instrument when selecting employees as it, for example, could examine the accordance with company values. Kim (2003) discovered that participants could successfully develop strategies to fake the results in an IAT when instructed to react slower. Remarkably, this was not done spontaneously. This can affect the argument for using implicit measures, such as the IAT or GNAT, when selecting employees, as they could improve their performance if they understood the underlying mechanism and did this test more than once (Kim, 2003). Asendorpf et al. (2002) raised concerns about the ethical use of implicit measures as they reveal participants' involuntary answers and are not under their own control. This data, therefore, needs to be handled carefully, and sufficient data security should be implemented.

Investors could use this measure to uncover the implicit attitudes and morality traits of founders or companies' top management (e.g., CEO). They could assess the morality of these managers and draw conclusions for their investment decisions. For example, investing only in honest (or moral) founder teams potentially implies a transparent and more honest environment, exchange of information, and prospective company success.

Policymakers could benefit from an effective tool to measure peoples' moral self-image and implicit attitudes to understand their concerns and reasonings. This could be applied to effectively nudge them toward more sustainable behavior by implementing reasonable restrictions and policies. Additionally, this could affect elections and politics, as voters feel better understood by politicians.

6. Conclusion

An effective tool to measure people's moral self-image provides valuable insights, extending the research in behavioral economics and moral balancing (Mazar & Zhong, 2010).

This thesis explored the effectiveness of the GNAT by examining the convergent validity between the GNAT and the explicit moral self-image questionnaire. It discovered that a lower correlation between implicit and explicit measures, thus a lower convergent validity, was obtained if a set of only 20 stimuli words were repeated four times in each block

(group A). This was explained by the participants' learning effects throughout the experiment, decreasing the convergent validity and effectiveness of the GNAT. For this reason, this thesis postulates using different stimuli without repetition instead, to avoid learning effects. This approach yielded a higher convergent validity (correlation) between implicit and explicit measures of the moral self-image, indicating higher effectiveness of the GNAT (group B).

These findings are crucial to effectively measuring the moral self-image, which is essential later on to investigate and understand its changes when aiming to predict moral behavior. This knowledge can be used for practical applications in behavioral economics, such as corporate management or policymaking.

Furthermore, this thesis contributes to research with its experimental design and analysis of the GNAT, including the adapted exclusion criteria, a sample correction (of all perfect response rates), and a first attempt to analyze the results based on reaction times (see Appendix V.IV Implicit Moral Self-Image Score Based on Reaction Times). This offers promising approaches for future research, which should particularly focus on a joint analysis of reaction times and response rates, experimental adjustments of the GNAT (e.g., changing the block order), and measuring the participant's moral self-image over a longer timeframe. Further research is necessary to validate the power of the GNAT to predict (e.g., consistent or balancing) behavior before bringing it into practical use in consumer behavior, for instance, by nudging people toward purchasing a more expensive but sustainable product.

However, the GNAT yields promising validity and insights into understanding people's implicit moral self-image and actions. This can be a powerful tool in the future, especially since it can be conducted entirely online, which helps assess the moral self-image in daily life (outside a laboratory environment).

Next time you order something online, use non-recyclable packaging, choose non-organic products, do not separate food waste appropriately, or book a plane ticket, remember the effect of these decisions on the environment and ultimately on your moral self-image, which influences future behavior.

References

- Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83(6), 1423–1440. <https://doi.org/10.1037/0022-3514.83.6.1423>
- Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology*, 83(2), 380–393. <https://doi.org/10.1037/0022-3514.83.2.380>
- Ashford, R. D., Brown, A. M., & Curtis, B. (2019). The Language of Substance Use and Recovery: Novel Use of the Go/No-Go Association Task to Measure Implicit Bias. *Health Communication*, 34(11), 1296–1302. <https://doi.org/10.1080/10410236.2018.1481709>
- Asuero, A. G., Sayago, A., & González, A. G. (2006). The Correlation Coefficient: An Overview. *Critical Reviews in Analytical Chemistry*, 36(1), 41–59. <https://doi.org/10.1080/10408340500526766>
- Barkan, R., Ayal, S., & Ariely, D. (2015). Ethical dissonance, justifications, and moral behavior. *Current Opinion in Psychology*, 6, 157–161. <https://doi.org/10.1016/j.copsyc.2015.08.001>
- Bartels, J. M., & Schoenrade, P. (2022). The Implicit Association Test in Introductory Psychology Textbooks: Blind Spot for Controversy. *Psychology Learning & Teaching*, 21(2), 113–125. <https://doi.org/10.1177/14757257211055200>
- Bassett, J. F., & Dabbs, J. M. (2005). A portable version of the go/no-go association task (GNAT). *Behavior Research Methods*, 37(3), 506–512. <https://doi.org/10.3758/BF03192721>
- Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, 76(2), 169–217. <https://doi.org/10.1086/259394>
- Blasi, A. (1993). The Development of Identity: Some Implications for Moral Functioning. In G. G. Noam & T. E. Wren (Eds.), *Studies in contemporary German social thought. Moral self: Working conference: Revised papers* (pp. 99–122). MIT Press.
- Boldero, J. M., Rawlings, D., & Haslam, N. (2007). Convergence between GNAT-assessed implicit and explicit personality. *European Journal of Personality*, 21(3), 341–358. <https://doi.org/10.1002/per.622>
- Brownstein, M., Madva, A., & Gawronski, B. (2020). Understanding Implicit Bias: Putting the Criticism into Perspective. *Pacific Philosophical Quarterly*, 101(2), 276–307. <https://doi.org/10.1111/papq.12302>
- Buhlmann, U., Teachman, B. A., & Kathmann, N. (2011). Evaluating implicit attractiveness beliefs in body dysmorphic disorder using the Go/No-go Association Task. *Journal of Behavior Therapy and Experimental Psychiatry*, 42(2), 192–197. <https://doi.org/10.1016/j.jbtep.2010.10.003>
- Carlson, K. D., & Herdman, A. O. (2012). Understanding the Impact of Convergent Validity on Research Results. *Organizational Research Methods*, 15(1), 17–32. <https://doi.org/10.1177/1094428110392383>
- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97. <https://doi.org/10.1016/j.jbef.2015.12.001>
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. <https://doi.org/10.4324/9780203771587>
- Cornelissen, G., Bashshur, M. R., Rode, J., & Le Menestrel, M. (2013). Rules or consequences? The role of ethical mind-sets in moral dynamics. *Psychological Science*, 24(4), 482–488. <https://doi.org/10.1177/0956797612457376>
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349–354. <https://doi.org/10.1037/h0047358>
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, 12(2), 163–170. <https://doi.org/10.1111/1467-9280.00328>
- Dabbs, J. M., Bassett, J. F., & Dyomina, N. V. (2003). The palm IAT: A portable version of the implicit association task. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, 35(1), 90–95. <https://doi.org/10.3758/BF03195500>
- de Houwer, J. (2006). What are Implicit Measures and Why are We Using Them? In R. Wiers & A. Stacy (Eds.), *Handbook of Implicit Cognition and Addiction* (pp. 11–28). SAGE Publications, Inc. <https://doi.org/10.4135/9781412976237.n2>
- de Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, 135(3), 347–368. <https://doi.org/10.1037/a0014211>
- de Houwer, J. (2001). A Structural and Process Analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, 37(6), 443–451. <https://doi.org/10.1006/jesp.2000.1464>
- Doherty, K., & Schlenker, B. R. (1991). Self-Consciousness and Strategic Self-Presentation. *Journal of Personality*, 59(1), 1–18. <https://doi.org/10.1111/j.1467-6494.1991.tb00765.x>
- Effron, D. A., & Monin, B. (2010). Letting people off the hook: When do good deeds excuse transgressions? *Personality & Social Psychology Bulletin*, 36(12), 1618–1634. <https://doi.org/10.1177/0146167210385922>

- Erat, S., & Gneezy, U. (2012). White Lies. *Management Science*, 58(4), 723–733. <https://doi.org/10.1287/mnsc.1110.1449>
- Erikson, E. H. (1964). *Insight and responsibility: Lectures on the ethical implications of psychoanalytic insight* ([Norton pbk. ed.]). W.W. Norton.
- Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., & Tutz, G. (2016). *Statistik*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-50372-0>
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition. Research: Their meaning and use. *Annual Review of Psychology*, 54, 297–327. <https://doi.org/10.1146/annurev.psych.54.10160.1.145225>
- Ferguson, R. (2018). *How flexible is morality? A test of the moral credits model of moral balancing* [Doctoral dissertation, Australian Catholic University].
- Festinger, L. (1957). *A theory of cognitive dissonance* ([Renewed 1985 by author]). Stanford University Press.
- Fischer, J., Dyball, R., Fazey, I., Gross, C., Dovers, S., Ehrlich, P. R., Brulle, R. J., Christensen, C., & Borden, R. J. (2012). Human behavior and sustainability. *Frontiers in Ecology and the Environment*, 10(3), 153–160. <https://doi.org/10.1890/110079>
- Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, 4(2), 195–202. <https://doi.org/10.1037/h0023552>
- Friese, M., Bluemke, M., & Wänke, M. (2007). Predicting voting behavior with implicit attitude measures: The 2002 German parliamentary election. *Experimental Psychology*, 54(4), 247–255. <https://doi.org/10.1027/1618-3169.54.4.247>
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley New York.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197–216. <https://doi.org/10.1037/0022-3514.85.2.197>
- Jacobsen, C., Fosgaard, T. R., & Pascual-Ezama, D. (2018). WHY DO WE LIE? A PRACTICAL GUIDE TO THE DISHONESTY LITERATURE. *Journal of Economic Surveys*, 32(2), 357–387. <https://doi.org/10.1111/joes.12204>
- Johnston, M. E., Sherman, A., & Grusec, J. E. (2013). Predicting moral outrage and religiosity with an implicit measure of moral identity. *Journal of Research in Personality*, 47(3), 209–217. <https://doi.org/10.1016/j.jrp.2013.01.006>
- Jordan, J., Leliveld, M. C., & Tenbrunsel, A. E. (2015). The Moral Self-Image Scale: Measuring and Understanding the Malleability of the Moral Self. *Frontiers in Psychology*, 6, 1878. <https://doi.org/10.3389/fpsyg.2015.01878>
- Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality & Social Psychology Bulletin*, 37(5), 701–713. <https://doi.org/10.1177/0146167211400208>
- Kadlec, H. (1999). Statistical properties of d' and β estimates of signal detection theory. *Psychological Methods*, 4(1), 22–43. <https://doi.org/10.1037/1082-989X.4.1.22>
- Kim, D. Y. (2003). Voluntary Controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly*, 66(1), 83. <https://doi.org/10.2307/3090143>
- Lee, Y. H., & Hsieh, G. (2013). Does slacktivism hurt activism? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 811–820. <https://doi.org/10.1145/2470654.2470770>
- Macmillan, N. A. (2002). Signal Detection Theory. In H. Pashler (Ed.), *Stevens' Handbook of Experimental Psychology: Volume 4: Methodology in Experimental Psychology. THIRD EDITION* (pp. 43–90). John Wiley & Sons, Inc. <https://doi.org/10.1002/0471214426.pas0402>
- Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, 45(6), 633–644. <https://doi.org/10.1509/jmkr.45.6.633>
- Mazar, N., & Zhong, C. B. (2010). Do green products make us better people? *Psychological Science*, 21(4), 494–498. <https://doi.org/10.1177/0956797610363538>
- Melé, D., & Cantón, C. G. (2014). The Homo Economicus Model. In *Human Foundations of Management* (pp. 9–29). Palgrave Macmillan UK. https://doi.org/10.1057/9781137462619_2
- Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral Self-Licensing: When Being Good Frees Us to Be Bad. *Social and Personality Psychology Compass*, 4(5), 344–357. <https://doi.org/10.1111/j.1751-9004.2010.00263.x>
- Miller, D. T., & Effron, D. A. (2010). Chapter Three - Psychological License: When it is Needed and How it Functions. In *Advances in Experimental Social Psychology* (pp. 115–155, Vol. 43). Academic Press. [https://doi.org/10.1016/S0065-2601\(10\)43003-8](https://doi.org/10.1016/S0065-2601(10)43003-8)
- Monin, B., & Jordan, A. H. (2009). The Dynamic Moral Self: A Social Psychological Perspective. In *Personality, Identity, and Character* (pp. 341–354). Cambridge University Press. <https://doi.org/10.1017/CBO9780511627125.016>
- Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81(1), 33–43. <https://doi.org/10.1037/0022-3514.81.1.33>
- Mullen, E., & Monin, B. (2016). Consistency Versus Licensing Effects of Past Moral Behavior. *Annual Review of Psychology*, 67, 363–385. <https://doi.org/10.1146/annurev-psych-010213-115120>
- Nisan, M. (1990). Moral Balance: A Model of How People Arrive at Moral Decisions. In T. E. Wren (Ed.), *The Moral Domain: Essays in the Ongoing Discussion between Philosophy and the Social Sciences* (pp. 283–314). The MIT Press.
- Nosek, B. A., & Banaji, M. R. (2001). The Go/No-Go Association Task. *Social Cognition*, 19(6), 625–666. <https://doi.org/10.1521/soco.19.6.625.20886>
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598–609. <https://doi.org/10.1037/0022-3514.46.3.598>
- Perkins, B. G., Podsakoff, N. P., & Welsh, D. T. (2024). Variance in Virtue: An Integrative Review of Intraindividual (Un)Ethical Behavior Research. *Academy of Management Annals*, 18(1), 210–250. <https://doi.org/10.5465/annals.2022.0057>
- Perugini, M. (2005). Predictive models of implicit and explicit attitudes. *The British Journal of Social Psychology*, 44(Pt 1), 29–45. <https://doi.org/10.1348/014466604X23491>
- Perugini, M., & Leone, L. (2009). Implicit self-concept and moral action. *Journal of Research in Personality*, 43(5), 747–754. <https://doi.org/10.1016/j.jrp.2009.03.015>
- Ploner, M., & Regner, T. (2013). Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization*, 93, 374–383. <https://doi.org/10.1016/j.jebo.2013.03.030>
- Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science*, 20(4), 523–528. <https://doi.org/10.1111/j.1467-9280.2009.02326.x>
- Sachdeva, S., Jordan, J., & Mazar, N. (2015). Green consumerism: moral motivations to a sustainable future. *Current Opinion in Psychology*, 6, 60–65. <https://doi.org/10.1016/j.copsy.2015.03.029>
- Schimmack, U. (2021). The Implicit Association Test: A Method in Search of a Construct. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 16(2), 396–414. <https://doi.org/10.1177/1745691619863798>
- Schlegelmilch, B. B., & Simbrunner, P. (2019). Moral licensing and moral cleansing applied to company-NGO collaborations in an online context. *Journal of Business Research*, 95, 544–552. <https://doi.org/10.1016/j.jbusres.2018.07.040>
- Schnabel, K., Asendorpf, J., & Greenwald, A. (2007). Using Implicit Association Tests for the Assessment of Implicit Personality Self-Concept. In G. J. Boyle, G. Matthews, & H. Saklofske (Eds.), *Handbook of Personality Theory and Testing*. SAGE Publications, Inc. <https://doi.org/10.4135/9781849200479.n24>
- Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-Serving Justifications. *Current Directions in Psychological Science*, 24(2), 125–130. <https://doi.org/10.1177/0963721414553264>

- Teachman, B. A. (2007). Evaluating implicit spider fear associations using the Go/No-go Association Task. *Journal of Behavior Therapy and Experimental Psychiatry*, 38(2), 156–167. <https://doi.org/10.1016/j.jbtep.2006.10.006>
- Williams, B. J., & Kaufmann, L. M. (2012). Reliability of the Go/No Go Association Task. *Journal of Experimental Social Psychology*, 48(4), 879–891. <https://doi.org/10.1016/j.jesp.2012.03.001>
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107(1), 101–126. <https://doi.org/10.1037/0033-295X.107.1.101>
- Zhong, C. B., Ku, G., Lount, R. B., & Murnighan, J. K. (2010). Compensatory Ethics. *Journal of Business Ethics*, 92(3), 323–339. <https://doi.org/10.1007/s10551-009-0161-6>
- Zhong, C. B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science (New York, N.Y.)*, 313(5792), 1451–1452. <https://doi.org/10.1126/science.1130726>



Who Bears the Costs of the UK Soft Drink Tax? An Empirical Study of Medium-Term Effects

Laura Wiredu

Paderborn University

Abstract

Using five years post-tax data on CPI prices, national employment as well as firm-level employment, we provide novel evidence on the medium-term effects of the UK Soft Drinks Industry Levy (SDIL). Applying a difference-in-differences research design, we find that neither consumers nor employees visibly bore the costs of the tax. However, given the falling trend in beverage prices in the UK, consumers faced a 6% less decline in soft drink prices compared to prices of levy-exempt beverages, since the introduction of the tax. Employees in the soft drink manufacturing industry were unaffected by the tax, whilst employees in the beverage manufacturing industry even benefitted from it through significant employment increases. Our study offers policy advice for other countries which are yet to implement soft drink taxes. We advocate the implementation of tiered tax designs, such as in the UK, due to its positive impact on employment and minimal burden for consumers.

Keywords: public health policy; soft drink taxation; UK sugar tax

1. Introduction

Increasing health concerns, such as obesity and the associated non-communicable diseases (NCDs)¹, have urged governments worldwide to address the excess sugar intake of the population (World Health Organization, 2023a). Hereby, the UK is one of the countries suffering the most from this health issue, having the third highest obesity rate in Europe in 2016, succeeding Malta and Turkey (Metcalf & Sasse, 2023). By 2016, it was inevitable for the UK government to encounter this adiposity epidemic, with an obesity rate striking nearly 30% (Metcalf & Sasse, 2023).² Consequently, the UK Soft

Drinks Industry Levy (SDIL)³ was implemented in April 2018 (HM Revenue & Customs, 2024a). Its two-tiered policy design aims at encouraging manufacturers to reduce the sugar content of their beverage portfolio through reformulation (HM Revenue & Customs, 2024a); the less sugar is added to a drink the lower the tax burden. However, as prevalent for excise taxes, producers often shift the extra costs arising through the tax onto other stakeholder groups to maintain pre-tax profits. Considering that the UK soft drink tax is a producer-imposed tax, it is not intended to be passed down to consumers (e.g. in the form of higher product prices) or employees (e.g. job cuts). Thus, to understand the effectiveness of the UK soft drink tax, it is vital to first question “Who bears the costs of the UK soft drink tax?”. To identify which party carried the tax burden I perform difference-in-differences analyses using firm-level and national employment data, as well as product price data on soft drink prices. I find that neither employees nor consumers visibly bear the tax burden. However, the tax introduction decelerated the

I would like to express my sincere appreciation to my supervisor, Dr. Henning Giese, whose invaluable input and continuous support made this publication possible. My deepest gratitude also goes to my family, friends and (former) managers, without whom I would not be where I am today. Their unwavering support and trust gave me the courage to pursue and achieve my academic goals.

¹ Non-communicable diseases are chronic diseases, such as cardiovascular health issues (e.g. strokes), diabetes, cancer etc. They account for 74% of worldwide deaths (World Health Organization, 2023b).

² As shown in Figure 2 in Metcalf and Sasse (2023).

³ For simplified comprehension, I hereafter speak of the UK soft drink tax in lieu of the UK Soft Drinks Industry Levy.

general trend of falling soft drink prices in the UK. By using medium-term data, my study captures effects which have not been materialised in other previous short-term studies on employment and prices.

This paper is structured as follows. Chapter 2 outlines the institutional setting of the UK soft drink tax and offers an extensive literature review on soft drink taxation. In Chapter 3, I outline my empirical strategy and describe the data. Chapter 4 elaborates on the empirical results. Robustness checks are presented in Chapter 5. Lastly, Chapter 6 concludes, summarises limitations and offers an outlook on future research.

2. Theoretical Background

2.1. Institutional Setting

Excise taxes have become a favourable tool to nudge the population's consumption pattern into a certain direction. Often referred to as "sin taxes", they are levied on the consumption of goods which are associated with (high) future costs for society (Allcott et al., 2019; de la Feria, 2024). For example, the intake of tobacco and alcohol or the use of fuel often come at the cost of the environment or a person's health. By taxing these types of products, their prices increase, leading to a decline in demand and herewith limiting their negative impacts (de la Feria, 2024).

Nowadays, excise taxes are often levied on sugar-sweetened beverages (SSBs),⁴ in order to curb the population's excess sugar intake. While the WHO's recommendation on an adult's daily sugar consumption is 50 g, a regular 600 ml soft drink bottle⁵ already exceeds this threshold (World Bank, 2020). Hence, worldwide obesity levels have been continuously rising since more than four decades (Metcalf & Sasse, 2023), making excise taxes on sugary drinks inevitable nowadays. According to the World Health Organization (2023a), more than 108 countries have adopted sugar-sweetened beverage taxes as at July 2022. Whilst the UK is one of the European countries most affected by high obesity rates (Metcalf & Sasse, 2023), it is also one of the most recent countries in Europe to have adopted a tax on sugar-sweetened beverages (see World Bank, 2020, Appendix 1).

Thus, the UK Soft Drinks Industry Levy, a national tax on sugar-sweetened beverages, was announced in March 2016 and implemented in April 2018 (HM Revenue & Customs, 2024a). The excise tax is imposed on soft drink producers and importers and comprises all types of soft drinks with more than 5 g of added sugar per 100 ml. Alcoholic drinks with more than 1.2% alcohol content, milk products, as well as fruit and vegetable juices are exempt from the levy (HM Revenue & Customs, 2024a). A tiered volume-based tax rate is applied to motivate producers to reduce the sugar content of their beverages by reformulating them. The two tiers are

defined as follows: 18 pence per litre for beverages with a sugar content between 5 g to 8 g per 100 ml; 24 pence per litre for beverages with more than 8 g sugar per 100 ml (HM Revenue & Customs, 2024a). In April 2023, the tax was extended to also encompass concentrates with more than 5 g sugar per 100 ml. All concentrates are taxed at the higher rate of 24 pence per litre (HM Revenue & Customs, 2024b).

Whilst a common consensus has revealed that sugar-sweetened beverage taxes can successfully reduce sugar consumption and herewith improve public health (World Bank, 2020), the estimated level of impact varies significantly across studies and countries. This can be partially attributed to the different policy designs and through which channels they are passed through. Whilst uniform taxes aim at cutting consumption levels through higher product prices for consumers, tiered tax designs focus on reducing overall sugar levels in beverages through product reformulation on the manufacturer's side (World Health Organization, 2023a). Despite most countries applying uniform tax rates (e.g. a single tax rate on all sugary products), the UK government decided to introduce a tiered tax rate to decrease the sugar content in soft drinks instead of just targeting consumption through prices. Thus, its primary objective is to incentivise soft drink manufacturers to reformulate their products and to introduce low-sugar alternatives in their product portfolio in order to evade the tax (HM Revenue & Customs, 2016). Therefore, the UK government provided manufacturers with a two-year lead time to improve their sugar-sweetened beverage portfolio beforehand. Hence, the change should already take place on the manufacturing side instead of just transferring the tax burden onto the consumer.

As obesity levels kept on increasing (Metcalf & Sasse, 2023), despite the introduction of the soft drink tax, the British government was forced to further intervene in the population's diet choices. Hence, in April 2022, the government implemented a compulsory calorie labelling law. All restaurants and other businesses (e.g. online deliveries) offering non-prepacked food (i.e. served meals) are obliged to display calorie information on their menu (Department of Health and Social Care, 2022). I assume that this law could affect employment levels in the soft drink industry in a negative way due to its signalling effect. Consumers might refrain from purchasing soft drinks in a restaurant or cafe, once they are reminded of the high amount of calories on the menu; the tax signals the customer how unhealthy the beverage is. Thus, SSB consumption might decrease in restaurants, leading to an overall lower demand for soft drink manufacturing. This continuous decline in demand could eventually lead to job losses in the respective industry. However, I argue that this law is unlikely to distort my results, as its introduction only overlaps with my data by one year (i.e. in 2023). I also perform a robustness check, in which the overlapping year is excluded, and find that my results remain robust (see Chapter 6).

⁴ SSBs are drinks which contain free sugars (i.e. added sugars and sugars naturally inherent in fruit juices and concentrates, syrups and honey) (World Health Organization, 2023a).

⁵ A 600 ml soft drink bottle approximately contains 64 g of sugar according to the World Bank's (2020) example.

2.2. Literature Review

Excise taxes entail additional costs for soft drink firms and could jeopardise their profits. Thus, firms often tend to pass the tax burden onto other stakeholder groups, such as employees (e.g. reductions in staffing) or consumers (e.g. increases in prices). Hereby, current literature extensively analyses price effects, as most sugar-sweetened beverage taxes are transmitted through this channel. Effects on employment remain an understudied area. Due to the novelty of most soft drink taxes, the vast majority of empirical studies are only able to capture the short-term effects of SSB taxation. Whilst this might be the case for price effects, I expect structural changes in employment to take longer to materialise.

For example, Scarborough et al. (2020) analyse the short-term effects on prices evoked by the UK soft drink tax. They conduct an interrupted time series analysis, using more than 210,000 observations on soft drink sugar content and prices. Primarily, they divide the sample into own branded products from the supermarket chain and famous branded products, as well as low-levy products, which have a sugar content between 5 g to 8 g per 100 ml and high-levy beverages with a sugar content of 8g per 100 ml or above. Firstly, they discover a reduction of the sugar content for both, “no name” and branded sugar-sweetened beverages. Secondly, they find that prices increased by 11.8 pence per litre for branded products from the higher levy tier, passing through approximately half of the tax. Surprisingly, the prices of the branded lower taxed beverages decreased by 17.4 pence per litre. For own branded drinks the opposite effect occurred. Prices of higher taxed drinks decreased by 63 pence per litre whereas prices of lower taxed products increased by 69 pence per litre (Scarborough et al., 2020).

Clearer industry responses have been observed in the Mexican and Chilean sugar sweetened beverage markets, after the implementation and increases of their national soft drink taxes. Cuadrado et al. (2020) conduct a time series analysis to regress the increase of the Chilean soft drink tax in 2014 on soft drink prices. Their analysis is based on consumer price data two years post tax implementation. The authors estimate a 5.6% price increase of carbonated beverages, which is equivalent to a 140% pass-through of the tax. Prices of untaxed beverages, such as juices, were unaffected. Contrary to the UK soft drink tax, the Chilean SSB tax was a consumer-imposed tax, aiming to reduce the affordability of soft drinks and to stimulate healthier substitution effects. As of that, the increase in soft drink prices has been the main intention of the Chilean soft drink tax policy, which can therefore be viewed as fruitful. In the case of Mexico, Aguilar et al. (2021) also identify a significant positive effect of the national SSB tax on prices of taxed beverages. Using household purchase data from Kantar World Panel, they conduct an event study. The authors estimate a 9.7% price increase of taxed soft drinks and a pass through rate on prices of 80%. Herewith, both tax designs exceed the WHO's recommendation on a minimal pass-through rate of 20% for SSB taxes to effectively reduce consumption (World Bank, 2020).

Another major study stems from Stacey et al. (2019), who analyse price effects in the frame of the South African SSB tax. The authors use Consumer Price Index (CPI) data to perform a before and after comparison one year post tax. They find that prices of carbonated beverages increased significantly, holding a pass-through rate of approximately 70%. Similar evidence on prices is also provided for the Saudi Arabian sugar-sweetened beverage tax. Alsukait et al. (2020) also use CPI price data to derive the price effects one year after Saudi Arabia's soft drink tax introduction. Their difference in differences analysis estimates significant price increases of 55% for carbonated beverages (which relates to a pass-through rate of 110%), compared to untaxed beverage prices.

As mentioned above, most analyses focus on the consumer tax incidence of sugar sweetened beverage policies. However, excise taxes can also be shifted onto employees, by decreasing wages or reducing the number of employees to maintain pre-tax profits. Most commonly, opponents of the tax argue that the UK soft drink tax leads to major job losses (Lauber et al., 2022) or a reduction in working hours (Cawley & Frisvold, 2023). They assume that the decline in soft drink demand harms the firm's economic performance and eventually leads to staff cuts (Hattersley et al., 2020). However, prevailing literature on employment does not validate these claims.

One of the main studies in this field stems from Marinello, Leider, Pugach, and Powell (2021), which analyses the effect of SSB taxes on employment in the US state Philadelphia. The authors conduct a pre-post analysis, based on employment data from the Bureau of Labor Statistics. Overall, they do not find a significant effect of the tax on the number of employees in the private sector. Similarly, Díaz et al. (2023) perform an interrupted time series analysis, using employment data from the Ministry of Labour in Peru. As they explore the effects of the national soft drink tax, they do not observe significant impacts on wage and employment levels in the beverage manufacturing sector (Díaz et al., 2023). Another study focussing on employment changes in the beverage manufacturing sector originates from Guerrero-López et al. (2017). Using monthly survey data on the soft drink and non-alcoholic beverage manufacturing industries, they find that the Mexican soft drink tax had no significant effect on employment in the respective industries. Furthermore, Gonçalves et al. (2024) provide recent evidence on employment effects in the setting of Portugal. As the Portuguese soda tax was implemented in 2017 and partially increased in 2019, the authors employ an event study design to derive short- and medium-term effects of the tax. They retrieve wage and employment data from annual tax declarations, covering a period from 2017 to 2019. The authors find that neither wages nor employment levels in SSB companies were affected.

Given the novelty of the UK soft drink tax, literature on its tax incidence is still quite scarce. While most studies on SSB taxes in other geographical settings imply price increases and no effect on employment, this might not be the case for the

UK due to its different policy design. Compared to other soft drink taxes, which aim at increasing prices to reduce consumption of sugary drinks, the British soft drink tax is not intended to be levied on consumers. For the tax to meet its primary target of reducing the population's sugar intake and herewith obesity levels, the sugar levels in soft drinks must be substantially decreased. This is underlined by the paper of Briggs et al. (2017), who argue that product reformulation has the greatest effect on the population's health in the UK. The authors conduct several modelling studies to quantify potential industry responses to the soft drink tax. They find that reducing the sugar content in drinks, by reformulating products, would have the greatest effect on public health. However, the effect partially diminishes if the industry decides to pass the tax onto the consumer, i.e. in the form of higher product prices (Briggs et al., 2017).

Hence, to analyse the effectiveness of the UK soft drink tax, it is vital to first question "Who bears the costs of the UK soft drink tax?". My research question herewith explicitly aims to identify whether the tax burden is borne by the company or shifted onto consumers or the firms' employees.⁶ The importance of identifying the incidence of the UK soft drink tax lies in the main objective of the tax. It was implemented to combat the UK's health pandemic of obesity by reducing sugar levels in drinks and herewith the overall sugar intake of the population (HM Revenue & Customs, 2016). Thus, it is intended to be primarily levied on the manufacturer side, to stimulate reformulation efforts through its two-tiered design (Scarborough et al., 2020). If passed onto consumers, in the form of higher product prices, the policy is less effective (Briggs et al., 2017). As reformulation can be costly and time intensive, manufacturers might not be necessarily encouraged to reduce the sugar content of their products and have the option to take the simpler approach of passing the tax down to consumers. A change in consumer choice, however, to purchase less sugary drinks, seems more difficult. Not having the option in the first place of purchasing so much sugar in soft drinks, as overall sugar levels have been reduced, seems more promising.

Thus, my approach focusses on identifying the incidence of the UK soft drink tax, as understanding industry reactions to the tax is essential to determine its efficiency (World Bank, 2020). Whilst a pass through of the tax onto consumers still has a positive, but reduced impact on the population's health (Briggs et al., 2017), I argue that a pass through onto employees does not affect public health at all and only bears negative consequences for society (e.g. less employment, lower wages). I postulate that companies can react in one of the following three ways to the tax. The first and most common option is to pass the tax onto consumers, by raising product prices. Another option for soft drink manufacturers is to

pass the tax burden onto their employees, by either reducing current staff or not increasing wages appropriately. This approach could be taken if firms want to avoid raising prices in order to remain competitive with other beverage manufacturers (e.g. juice manufacturers, which are not affected by the tax). Lastly, firms could avoid passing the tax burden onto other stakeholders due to market competitiveness and company reputation (e.g. reducing staff could stimulate negative publicity), but also circumvent the tax burden if they reformulate their soft drinks to fall below the sugar threshold. This response would be the UK government's preferred reaction to the tax, because reformulation leads to overall lower sugar levels in soft drinks. Hence, to identify how the British soft drink manufacturing industry reacted to the tax and who now bears the consequences of it (i.e. the tax burden), I postulate the subsequent hypotheses:

H1: The UK soft drink tax leads to a decrease in employment in the soft drink manufacturing industry.

H2: The UK soft drink tax leads to an increase in prices of soft drinks.

The first hypothesis aims to identify if consumers bear the cost of the tax due to rising soft drink prices. As elaborated above, the simplest approach for manufacturers to avoid the tax burden is to solely increase soft drink prices. In line with Cawley and Frisvold (2023), I argue that the pass through of the tax highly depends on the elasticity of demand and supply⁷ in a perfectly competitive market. If demand is inelastic, as there are no alternative products to the taxed soft drinks (i.e. no replacement product) and consumers are highly dependent on the good, the industry can fully pass the tax burden onto consumers (100% price pass through). Thus, despite the price increases, demand for soft drink products will not decline due to the consumers' dependency on that specific product. However, in the scenario of perfectly elastic demand, meaning there are many substitute products to the taxed soft drinks on the market, the consumer will opt for more cost-efficient alternatives, if prices of SSBs are increased. Thus, firms will refrain from raising prices and will fully absorb the tax burden (0% price pass through) or even offer price reductions, in order to stay competitive in the market. In line with Cawley and Frisvold (2023), I assume that the industry response will be somewhere in between the first scenario of fully passing the tax onto consumers and the second scenario of totally absorbing the tax burden. Similarly, the reaction of cutting staff (Hypothesis 2) seems intuitive, as increasing soft drink prices leads to a lower demand for the product. This leads to a decline in soft drink manufacturing and thus, eventually reduces the number of

⁶ I assume that shareholders will not significantly be impacted by the UK soft drink tax. This is because they react extremely sensitive to dividend pay-cuts, wherefore firms usually avoid this measure (Heiden, 2002) and only use it as a last resort. Thus, I argue that it is very unlikely that the soft drink tax burden is a valid reason for soft drink companies to reduce dividend payouts.

⁷ Elasticity of supply refers to the change of product supply due to a price increase. The change in the demand for a product (e.g. higher or lower consumption of it) due to price increases/decreases is known as the elasticity of demand (World Bank, 2020).

jobs in the soft drink manufacturing industry (Hattersley et al., 2020). This has been a typical argument put forward by the opposition to global soft drink taxes (Hattersley et al., 2020), which I deem to be plausible for the UK soft drink tax too. Alternatively, a possible counter argument to my second hypothesis is that consumers might substitute the now more expensive soft drinks with untaxed beverages such as low-calorie drinks or water (World Bank, 2020). Since soft drink manufacturers often produce a wide range of beverages (e.g. Coca-Cola also produces levy-exempt products such as water and juices), the substitution of products does not necessarily result in any employment losses. Rather, soft drink manufacturers might even require additional labour for reformulation and innovation of new low-sugar alternatives.

As to my knowledge, there has been no other empirical study explicitly analysing the UK soft drink tax effects on employment. As the few studies on the UK soft drink tax price effects provide twofold evidence, my analysis offers additional insights in this field. By evaluating employment and price effects within a medium-term horizon, my study accounts for structural changes that take longer to manifest in the market (e.g. structural shifts in employment and long-term manifested pricing strategies).

3. Empirical Method

3.1. Model and Methodology

To answer my research question of who bears the costs of the UK soft drink tax, I take three different approaches. First, I examine employment from a microeconomic scope, using firm-level data. Then I analyse the effect on a macroeconomic level, using national aggregate employment data of the beverage and soft drink manufacturing industries. Finally, I estimate price effects using product price data of soft drinks. I use two different employment approaches for more robust results. While the microeconomic data encompasses the most meaningful companies in the soft drink industry based on their market capitalisation and defined by The Refinitiv Business Data Classification, the macroeconomic employment analyses include randomly selected soft drink manufacturers regardless of their size. See Section 3.2 for a detailed description of the data, including its limitations.

I apply the same econometric framework to all analyses, namely difference-in-differences (DiD). This research design is common in policy analysis, as it compares the different outcomes of a treatment and control (comparison) group over time (i.e. before and after a new policy intervention) (Marcus & Sant'Anna, 2021). For DiD to identify the actual causal effect of the policy, the following prerequisite must be fulfilled. The treatment group should have developed in a similar way as the control group if no policy (treatment) had been implemented. This is also referred to as the parallel trends assumption (Angrist & Pischke, 2009; Marcus & Sant'Anna, 2021). Put in other words, both groups must have the same pre-intervention employment/price trends. To test this assumption, I make graphical comparisons and conduct event

studies. Event studies are commonly used for assessing the parallel trends assumption, as they offer a more granular view on the treatment effect (Marcus & Sant'Anna, 2021). Compared to DiD, which analyses the average change in outcomes (between treatment and control group) before and after the event, event studies look at the effect of the treatment at each specific period.

To estimate the causal effect of the soft drink tax on employment and price levels, I perform high-dimensional fixed effects regressions for all three specifications. Fixed-effects models account for unobserved variables, which are constant over time but differ across groups and herewith may affect the endogenous variable (Angrist & Pischke, 2009; Wooldridge, 2021). For example, the location of a company could influence the availability of labour due to the infrastructure and economic conditions in that area. My fixed effects regression model controls for this possible variation within units. As stated by Lee (2016), the fixed effects estimator is the preferred estimator for difference in differences models.

3.1.1. Microeconomic Employment Analysis

To test my first hypothesis, that is, whether the UK soft drink tax leads to a decline in employment in the UK soft drink manufacturing industry, I estimate the following model:

$$\begin{aligned} \text{Log}(\text{Employees})_{i,t} = & \beta_0 + \beta_1 \cdot \text{Tax}_t \\ & + \beta_2 \cdot \text{Treatment}_i \\ & + \beta_3 \cdot (\text{Tax}_t * \text{Treatment}_i) \\ & + \beta_k \cdot X_{i,t} + Y_{i,t-1} + \gamma_p + \varepsilon_{i,t} \end{aligned} \quad (1)$$

My dependent variable, $\text{Log}(\text{Employees})_{i,t}$ reflects the logarithmic number of employees in each company (i) per year (t). The variable Treatment_i is an indicator variable, taking the value one if a company is operating in the soft drink industry and zero if companies are alcoholic beverage producers. Tax_t is a binary variable, which is equal to zero for periods before the soft drink tax came into effect (2013-2017) and equal to one for years after the tax (2018 – 2023). My key variable of interest is the interaction term $\text{Tax}_t * \text{Treatment}_i$. It is also referred to as the treatment indicator, which takes the value of one if a unit is treated and situated in a period after the tax intervention. Otherwise, it takes the value of zero. Its coefficient β_3 , also known as difference-in-differences estimator, captures the causal effect of the UK soft drink tax on employment. As I use a log-linear specification, the coefficient estimates the difference in employment changes between soft drink firms and alcohol producing firms before and after the tax intervention. To control for possible confounders, the *Unemployment Rate*, *GDP* and *Total Population* are captured in the term $X_{i,t}$. I believe that these common macroeconomic indicators are likely to influence current employment levels in the soft drink manufacturing industry. Furthermore, I also control for company fixed effects (γ_p), such as the company's location for

example. All other potential confounders are encompassed in the error term $\varepsilon_{i;t}$.

Furthermore, my model also accounts for autocorrelation by including the first lag of the dependent variable ($Y_{i;t-1}$). Autocorrelation exists when the error term correlates across different time periods (Doran et al., 2017). In specific, when a variable can be defined by its previous year's value, this is called an autoregressive process (Das, 2019).⁸ Put in other words, I argue that the number of employees is strongly influenced by the number of employees in the preceding year.⁹ This could be caused by a company's hiring patterns or its economic growth over the past years. As I use annual data, it is sufficient to only control for the first lag (Doran et al., 2017).

Moreover, difference-in-differences models are often prone to cluster problems (Lee, 2016). This happens when error terms are correlated within a group (e.g. company) or across time periods (Imbens & Wooldridge, 2009). For example, companies often adapt their firm-specific hiring policies, which influences their number of employees over time. I control for this possible confounder by using cluster-robust standard errors,¹⁰ as suggested by Cameron and Miller (2015). In line with Gonçalves et al. (2024), I therefore cluster at the company level.

3.1.2. Macroeconomic Employment Analyses

In addition to the microeconomic analysis (i.e. firm-level employment), I perform two macroeconomic analyses (i.e. aggregate employment) for more robust results. I first regress the soft drink tax on employment in the soft drink manufacturing industry and afterwards on employment in the entire beverage manufacturing industry. Thus, I estimate the subsequent regression model to derive the macroeconomic employment effects evoked by the UK soft drink tax:

$$\begin{aligned} \text{Log}(\text{Employees})_{i;t} = & \beta_0 + \beta_1 \cdot \text{Tax}_t \\ & + \beta_2 \cdot \text{Treatment}_i \\ & + \beta_3 \cdot (\text{Tax}_t * \text{Treatment}_i) \\ & + \beta_k \cdot X_{i;t} + Y_{i;t-1} + \varphi_q + \varepsilon_{i;t} \end{aligned} \quad (2)$$

The dependent variable $\text{Log}(\text{Employees})_{i;t}$ reflects either the logarithmic number of employees in the beverage manufacturing industry or in the soft drink manufacturing industry. Equivalent to Equation (1), the term $X_{i;t}$ comprises the *Unemployment Rate*, *GDP* and *Total Population*.

⁸ When the variable is only influenced by its last year's value, this is referred to as an autoregressive process of order one or just AR(1) (Das, 2019).

⁹ I perform the Wooldridge test for serial correlation and find that my model suffers from autocorrelation issues. Furthermore, I also perform the Breusch-Godfrey test for autocorrelation, which tests if autoregressive processes larger than AR(1) (first lag) are present in the model (Doran et al., 2017). The test statistic indicates that my model does not suffer from autocorrelation beyond the first autoregressive process (AR(1)).

¹⁰ Cluster-robust standard errors control for both, cluster problems and for heteroskedasticity (Cameron & Miller, 2015).

Again, $Y_{i;t-1}$ denotes the lagged dependent variable¹¹ and $\varepsilon_{i;t}$ the error term. I further control for industry-specific fixed effects (φ_q).

Furthermore, I use heteroskedasticity-robust standard errors, but do not cluster at group levels due to an insufficient number of clusters. Not addressing possible cluster issues within the model could lead to an underestimation of standard errors (Lee, 2016). However, if the number of clusters is too small, this could also lead to underestimated standard errors (Cameron & Miller, 2015; Cawley & Frisvold, 2017). While Cameron and Miller (2015) postulate that more clusters lead to better statistical inference, their example underlines that fewer than 10 clusters result in extremely high rejection rates of the null hypothesis. I therefore refrain from using only two industry clusters. Thus, regression results should be treated with caution.

3.1.3. Price Analysis

Using the regression model below, I examine if the UK soft drink tax leads to increases in soft drink prices (Hypothesis 2):

$$\begin{aligned} \text{Log}(\text{Price})_{i;t} = & \beta_0 + \beta_1 \cdot \text{Tax}_t + \beta_2 \cdot \text{Treatment}_i \\ & + \beta_3 \cdot (\text{Tax}_t * \text{Treatment}_i) \\ & + \beta_k \cdot X_{i;t} + \omega_r + \delta_s + \varepsilon_{i;t} \end{aligned} \quad (3)$$

My dependent variable $\text{Log}(\text{Price})_{i;t}$ denotes the logarithmic real (inflation-adjusted) price per product (i) per year (t). In the above model specification, the term $X_{i;t}$ encompasses the control variables *Unemployment Rate*, *GDP*, *Total Population* and *Volume*. I argue that the first three variables can influence price levels, as these macroeconomic indicators reflect the health of the economy. If for example, there is an economic downturn, reflected in a high unemployment rate or low GDP, people tend to purchase less, leading to a decline in demand and finally decreasing prices. With the variable *Volume* I also control for the effect of the product size on prices, because small sized drinks tend to be more expensive than larger volume packs (World Health Organization, 2023a). Furthermore, I also control for region (ω_r) and shop type (δ_s) fixed effects. All the other variables (*Tax*, *Treatment*, *Tax * Treatment*, error term) are identical to Equation (1). Additionally, I cluster standard errors at shop level, to control for correlated outcomes within the same shop. Unfortunately, I was not able to include a lagged price variable in this model due to the missing panel structure of my data (see subsequent Section 3.2.3).

3.2. Data

For a unified comparison, I standardise time periods and nearly all control variables for all three analyses.¹² All

¹¹ In line with the previous microeconomic model (see Equation (1)), I perform the same tests for autocorrelation and thus, include a lagged dependent variable.

¹² Besides the uniform control variables *Unemployment Rate*, *GDP* and *Total Population* I additionally control for *Volume* in the price analysis.

datasets cover periods from 2013 to 2023, ensuring balanced pre- and post-tax periods. I obtained my control variables *GDP* and *Total Population* from the World Bank website. The variable *GDP* reflects the UK's annual Gross Domestic Product per capita in current US \$ (World Bank, 2023a). *Total Population* encompasses all residents in the UK independent of their legal status in the country (World Bank, 2023b). I obtain the annual *Unemployment Rate* from the Office for National Statistics. The variable depicts the proportion of unemployed individuals aged 16 years or older in relation to the working population (Office for National Statistics, 2020). Missing values of the dependent variables and controls are excluded from all analyses. Detailed summary statistics and descriptions of all variables can be found in the Annex in Tables A1 and A2.

3.2.1. Microeconomic Employment Data

I obtain annual employee data from Refinitiv Eikon Worldscope, a global database which provides financial and industry information on public companies. Refinitiv is renowned for its trusted global equity indices used for market analysis. Their indices capture significant firms in a specific stock market, by including 99.5% of all publicly traded shares in a (trading) market (Refinitiv, 2023). Their sector indices are built upon the global equity indices, by appointing companies into pre-defined industry groups according to the TRBC. The Reference Data Business Classification (TRBC) is a categorisation system which groups companies in accordance with their business operations. The TRBC divides the beverage industry group into the following industries: brewers, distillers & wineries and non-alcoholic beverages. As the latter definition is quite broad, the database also offers more granular information on the company's main economic activity with the NAICS code (North American Industry Classification System).

I use the Refinitiv sector index of the UK beverage industry to derive a comprehensive constituent list of 15 UK beverage manufacturers. However, I restrict my analysis to companies having their main sales market in the UK, due to the subsequent reasons.¹³ As I hypothesize that soft drink companies shift the tax burden onto their employees, I am focussing on companies directly affected by the soft drink tax. Hence, I postulate that beverage manufacturers must have their main sales market in the UK, for a portion of their sales to be noticeably affected by the new tax burden. Herewith, I exclude three companies¹⁴ from the initial constituent list.

¹³ My definition of main sales market is the following: either the United Kingdom (UK) or Great Britain (GB) must be within a company's three biggest sales markets according to their annual report in 2023.

¹⁴ I exclude East Imperial as the EU is only its fifth biggest sales market, accounting for 10% of their annual sales in 2022 (London Stock Exchange, 2023). Rogue Baron is excluded from the analysis, as no employment data was available in Worldscope. Momentous Holdings Corp was also excluded, as it is a holding company (not directly involved in beverage manufacturing itself), which currently does not engage in any business operations, has no revenue (Reuters, n.d.) and does not have any publicly available information.

See Table 1 for a detailed constituent list and information on the companies' UK market shares.

Table 1: Microeconomic Employment Data

#	Treatment Group	Market Cap. in £b	UK Sales in %
1	Coca-Cola Europacific Partners	26.587	17.5% (GB)
2	Britvic	2.406	71,4%
3	Fever-Tree Drinks	1.193	32%
4	A G Barr	0.675	95.6%
5	Nichols	0.365	74%
Control Group			
1	Diageo	57.633	9.1 % (GB)
2	Chapel Down Group	0.114	97.8%
3	Daniel Thawties	0.044	100%
4	Gusbourne	0.039	78.8%
5	Artisinal Spirits Company	0.027	29.8%
6	Adnams	0.006	98%
7	Distil	0.004	90.2%

Note. Author's own overview based on market capitalisation data retrieved from Refinitiv (2024a) and sales data retrieved from annual reports in 2023 of the respective companies. The graph depicts 12 of the 15 constituents of the TRBC beverage industry group, which are included in my regression analysis. Market capitalisation is depicted in billion pounds.

Based on the TRBC of the UK beverage sector, I appoint companies to the treatment and control group as follows. Enterprises operating in the non-alcoholic beverage market are assigned to the treatment group,¹⁵ and hereafter referred to as soft drink companies, as their main economic activity lies in soft drink manufacturing according to their NAICS code. The control group contains companies from the distillers & wineries and brewers industries and are hereafter referred to as alcoholic beverage companies.

My final sample consists of 12 beverage manufacturers, resulting in a total of 106 firm-year observations. I have an unbalanced panel, as not all companies reported employee data for each year. My dependent variable of interest, the number of employees, comprises both, full and part-time employees in a company (Refinitiv, 2024b). The number of employees range between seven to thirty thousand. In line with Guerrero-López et al. (2017), I use the logarithmic number of employees to control for the right skewness (no normal distribution) in the dependent variable. Due to the high variance in the number of employees, using the logarithm also makes

¹⁵ Contrary to the TRBC classification, I appoint Fever-Tree Drinks to the treatment group, as its main product segment are carbonated non-alcoholic mixers, which are liable to the UK soft drink tax (Young, 2016).

the final results more comparable, as they will be interpreted as percentage changes instead of absolute values.

One limitation of my data is that the sector indices only cover publicly listed companies. However, Figure 1 indicates that the UK beverage market is dominated by a few, very large manufacturers. The graph depicts the UK's major beverage companies in terms of their market capitalisation as at October 2024 (Statista, 2024). It indicates that Diageo, Coca-Cola Europacific Partners and Britvic have a substantial lead in terms of market shares. I argue that manufacturers which are not listed on the stock exchange, are usually insignificant in size with a low number of employees, compared to the leading beverage manufacturers depicted in Figure 1. Furthermore, I partially account for these smaller manufacturers in my macroeconomic employment analyses (see Section 3.2.2). Hence, both micro- and macroeconomic employment analyses are representative for employment in the UK beverage sector.

3.2.2. Macroeconomic Employment Data

To analyse the effect of the soft drink tax on macroeconomic employment, I retrieve annual employment data per industry from the Business Register and Employment Survey (BRES). This annual survey covers 85,000 randomly selected companies registered for VAT and PAYE¹⁶ in Great Britain and is considered as the government's predominant source for employment estimates (Office for National Statistics, 2022a). In the survey industries are categorised according to the UK Standard Industrial Classification (SIC). This system appoints companies into pre-defined industries based on their primary economic activity. SIC codes offer a five-level breakdown, from the section level (one digit) to the subclass level (five digits) (Office for National Statistics, 2009). I derive my statistics from the division level (e.g. manufacturer of beverages) and the class level (e.g. manufacturer of soft drinks, production of mineral waters and other bottled waters).

Based on this classification system, I define my treatment and control groups as follows. For my first macroeconomic analysis, I use manufacturers of soft drinks, production of mineral waters and other bottled waters as a treatment group. My control group consists of six alcoholic beverage manufacturing industries, which I argue to be not significantly affected by the tax reform, as beverages with an alcohol content of 1.2% or more are exempt from the tax. For my second broader macroeconomic analysis, the beverage industry serves as the treatment and the textile industry as the control group. I argue that the textile industry is unlikely to be affected by the soft drink tax and therefore offers a robust comparison group. It also fulfils the parallel trends assumption (see Chapter 4), as it exhibits the same pre-tax trend as the beverage industry before the reform in 2018. A constituent list of the treatment and control groups for my macroeconomic analyses can be derived from Table 2.

¹⁶ Pay as You Earn (PAYE) is the UK income tax system, which directly deducts the tax from the employees' wages (HM Revenue & Customs, n.d.).

For the soft drink manufacturing analysis, I have a balanced dataset consisting of 22 industry-year observations over the period 2013 to 2023. The sample for the analysis of the beverage industry contains 71 industry-year observations and is an unbalanced panel. For both analyses, the dependent variable of interest is the logarithmic number of employees, which encompasses full- and part-time employees in an industry (Office for National Statistics, 2022a).

A limitation of this study is that manufacturing data is only available for Great Britain and not the entire United Kingdom. A separate Business Register and Employment Survey is conducted for enterprises situated in Northern Ireland and partially provides employment figures per industry as of 2016. However, as this survey does not consistently report employment statistics for the sub-classes of the beverage industry (SIC codes 1101-1107), I exclude these statistics from my analysis and continue with using Great Britain's employment data only. I postulate that my estimates are representative for the complete UK beverage manufacturing industry, as only a small portion of the manufacturers are situated in Northern Ireland (3.6%). Table 3 underlines that my GB data represents a significant portion of the UK soft drink manufacturing industry (96.4%). Secondly, BRES employment data has an inconsistent definition. Data from 2009 to 2015 only surveys VAT-registered businesses in GB, whereas data as of 2016 onwards additionally includes PAYE businesses. Furthermore, I acknowledge the fact that my data does not distinguish between employees in soft drink, mineral water and bottled water manufacturing.

3.2.3. Price data

I retrieve annual product price data (CPI price quotes in January of each year) from the Office for National Statistics (ONS) website (Office for National Statistics, 2024b). ONS publishes the constituent list of the Consumer Price Index (CPI), namely the CPI price quotes, on a monthly basis.¹⁷ The CPI includes approximately 700 different products, for which prices are collected from roughly 20,000 different shops across the UK. Hereby, the CPI depicts a representative sample of consumer goods. If consumption patterns or product availability in the shop changes, some products get replaced. Prices are hand collected from randomly selected shops (Office for National Statistics, 2023). Furthermore, information on the region, product volume, shop code (i.e. which store the price was collected from) and shop type (i.e. if the price was collected from chains with more than 10 shops or from independent stores) can also be obtained from the dataset.

Moreover, current prices today (nominal price values) must be made comparable over the years due to rising inflation and changing purchasing power. To convert nominal prices into real (inflation-adjusted) prices, the nominal values are divided by an inflation adjustment factor, for example

¹⁷ The data is not badged as National Statistics and is only made available for research purposes (Office for National Statistics, 2024b).

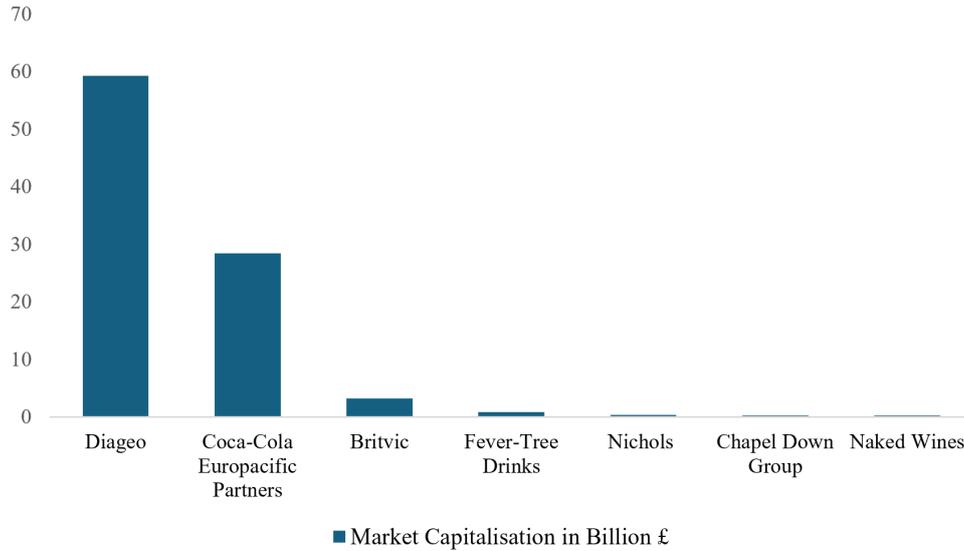


Figure 1: UK Leading Beverage Companies in 2024

Note. Adapted from “Leading beverage companies by market capitalization in the United Kingdom (UK) as of October 15,” by Statista, 2024 (<https://www.statista.com/statistics/1345274/uk-largest-beverage-companies-by-market-capitalization/>). Copyright 2024 by Statista. The original unit of the graph was converted from USD to GBP using the spot exchange rate of USD against GBP of 0.7888 on 15.10.2024 (Bank of England, n.d.). Despite that Naked Wines is listed as the seventh largest beverage manufacturer by Statista (2024), I have not included the company in my analysis. As Naked Wines is solely an online retailer of wine products and not a manufacturer, it is not listed in the Refinitiv sector index of the UK beverage industry, which I use to derive the key players of the UK beverage manufacturing industry.

Table 2: Macroeconomic Employment Data

#	Treatment Group	SIC Code	#	Control Group	SIC Code
1	Manufacturer of soft drinks, production of mineral waters and other bottled waters	1107	1	Distilling, rectifying and blending of spirits	1101
				Manufacture of wine from grape	1102
				Manufacture of cider and other fruit wines	1103
				Manufacture of other non-distilled fermented beverages	1104
				Manufacture of beer	1105
				Manufacture of malt	1106
2	Beverage Industry	11	2	Textile Industry	13

Note. Author’s own overview. Table depicts the underlying constituent list of the macroeconomic analyses. The first row shows the industries used in the soft drink industry analysis and the second row the industries used in the beverage industry analysis. The SIC code is the unique identifier for each industry, which I have used to derive the relevant industries from the BRES survey.

the CPI (Kumaranayake, 2000). I account for inflation by deflating all prices to 2018 (midpoint of my data) price levels. Thus, the prices in this paper are presented in 2018 constant pounds (£ 2018).¹⁸ I also standardise prices in terms of their product volumes, by using per litre prices in this study. The control variable *Volume* reflects the original volume in litres the product was sold in when its price was collected.

For my annual dataset I retrieve the CPI price quotes for January each year. I identify relevant beverage categories based on their COICOP classification. COICOP, which stands for the Classification of Individual Consumption According to Purpose, is a global framework used for classifying goods and services of household consumption (United Nations, 2018). The classification offers a granular four-level breakdown, from the two-digit division level (e.g. 01 = food and non-alcoholic beverages) to the five-digit subclass

¹⁸ This is equivalent to saying inflation-adjusted prices or real prices.

Table 3: Soft Drink Manufacturing Industry Employment Share in 2022

Geographic Region	Number of employees in soft drink, mineral water and bottled water manufacturing	% of UK Total
Great Britain	13,000	96.4%
Northern Ireland	490	3.6%
UK	13,490	100%

Note. Author's own overview based on employment data retrieved from the Business Register and Employment Survey published by Northern Ireland Statistics and Research Agency (NISRA) (NISRA, 2023) as well as the Business Register and Employment Survey published by the Office for National Statistics (2022b). The table shows the share of employees in the soft drink, mineral water and bottled water manufacturing industry across the UK's geographic regions in 2022.

level (e.g. 01.2.6.0 = soft drinks). I use the five-digit subclass level to allocate products to the treatment and control groups.¹⁹ All beverages under the soft drink classification are assigned to the treatment group. In line with the approaches of Gonçalves et al. (2024) and Scarborough et al. (2020), my control group is comprised of levy-exempt beverages, such as bottled water and fruit juices. An overview of the constituent list of the treatment and control group can be found in Table 4. As some items were substituted by other products over time,²⁰ I merged the observations for both goods to maintain a balanced time series (see shaded rows in Table 4). In 2019 the item cola flavoured drink (item ID 212011) was replaced by two different items (cola drink regular bottle and cola drink diet) of which one is taxable and the other non-taxable. To control for this discontinuity, I exclude all three items in my robustness checks and find that my estimates are robust to this change (see Chapter 5).

My dependent variable reflects the logarithmic real price per litre measured in pounds, in January each year. I exclude all prices equal to zero from my analysis, as this value indicates that the product was either temporarily out of stock or missing (i.e. not sold at the shop anymore) (Office for National Statistics, 2018). In line with Scarborough et al. (2020), I use the logarithmic price to account for the right skewness in the variable. My final sample comprises 31,957 price observations. Each item (e.g. Lemonade) has approximately 200 price observations per year, which each reflects a different hand collected price from a different store. Put differently, my price data has repeated cross-sections.

This missing panel structure leads to one of the major limitations in my analysis. Similar to Scarborough et al. (2020), I was not able to control for autocorrelation due to a missing unique product identifier. I construct a pseudo panel dataset,

by calculating the average price of each item ID per year. I find that my regression results are robust to this change (see Chapter 5). Additionally, due to the unavailability of nutritional information, I can not fully ensure that all beverages in the control group are unaffected by the soft drink tax. For example, flavoured water bottles containing more than 5 g of sugar per 100 ml are not exempt from the levy and should not be appointed to the control group. Furthermore, the missing information on sugar content does not allow me to differentiate between low-levy drinks and high-levy drinks. For example, the higher tax rate of 24 pence per litre (applicable to soft drinks with more than 8 g sugar per 100 ml) might have a more significant effect on prices than the lower tax rate of 18 pence per litre (applicable to soft drinks with 5 g to 8 g sugar per 100 ml).

4. Results

The results of my difference-in-differences analyses are reported separately in the following sub-chapters. As mentioned above, a prerequisite for difference-in-differences analysis to estimate the true causal impact of a policy is the parallel trends assumption. That is, both groups should follow a similar trend before the reform has been implemented (Angrist & Pischke, 2009; Marcus & Sant'Anna, 2021). This ensures that the difference in outcomes (of treatment and control group) after the reform, can be attributed to the policy, by controlling for the normal trend in both groups (e.g. employment/price trends) (see Lee, 2016). To test the parallel trends assumption, I have performed visual inspections of the pre-tax trends for the treatment and control groups (Figures 2, 4 and 6) and conducted separate event studies (Figures 1, 3 and 5) for each analysis below.

The graphical illustrations (Figures 1, 3 and 5) are all arranged in the following way. I normalised all values of the dependent variables to one in 2017, i.e. one period before the policy intervention, as employment levels and prices vary significantly across companies, industries and products. The vertical solid black line visualises the year when the UK soft drink tax came into effect (2018), whilst the grey dashed line depicts the year of its announcement (2016). The x-axis shows how the average value of the dependent variable

¹⁹ All items in the CPI price quote data have a four-digit class level (COICOP4 ID), but not every item is assigned a five-digit subclass level (COICOP5 ID). Information on the subclass level of item ID 310306 (bottle of mixer) is missing. As the item can be found in the overall class level mineral waters, soft drinks and juices, I still include it in the treatment group because sugar levels for mixers are usually higher than 5 g per 100 ml.

²⁰ Information on product replacements of the CPI constituents can be retrieved from the Office for National Statistics (2009, 2024a).

Table 4: Price Data

#	Treatment Group	Item ID	Start	End
1	Lemonade Bottle (2 L)	212008	2013	2023
2	Cola Flavoured Drink Bottle (1.75 - 2 L)	212011	2013	2019
3	Fizzy Drink / Cola Cans Pack 6-8 (0.33 L)	212012	2013	2016
4	Fizzy Energy Drink (0.25 – 0.5 L)	212015	2013	2023
5	Fizzy Bottled Drink (0.5 L)	212017	2013	2023
6	Mixer Drink Bottle (1 L)	212022	2015	2023
7	Cola / Fizzy Drink Pack 4-8 (0.33 L)	212023	2016	2023
8	Cola Drink Regular Bottle (1.25 – 2 L)	212025	2020	2023
9	Bottle of Mixer (0.125 – 0.2 L)	310306	2013	2023
Control Group				
1	Fruit Squash (0.75 - 1.5 L)	212001	2013	2023
2	Fruit Juice not Orange (1 L)	212006	2013	2023
3	Mineral Water Still (1.5 – 2 L)	212010	2013	2023
4	Fresh / Chill Orange Juice (0.9 L)	212016	2013	2023
5	Pure Fruit Smoothie (0.75 – 1 L)	212019	2013	2021
6	Fruit Drink Bottle 4-8 Pack (0.2 L)	212020	2013	2023
7	Bottle Still Water (0.5 L)	212021	2013	2023
8	Flavoured Water Bottle (0.9 - 1.5 L)	212024	2018	2023
9	Cola Drink Diet (1.25 - 2 L)	212026	2020	2023
10	Fruit / Vegetable Smoothie (0.75 – 1 L)	212027	2021	2023

Note. Author's own overview based on item descriptions from the CPI price quote data files (Office for National Statistics, 2024b). The table depicts the underlying items of my price analysis. All item IDs which were replaced over time are highlighted, as well as their replacement product. I assign item ID 212001 (fruit squash) to the control group, even though squashes are usually syrup based and need to be diluted with water. However, this drink would only fall under the UK soft drink tax as of April 2023 (extension of the tax to also include concentrates). As the last observations in my analysis are from January 2023, and squashes were taxable after April 2023, the item was tax-exempt during my entire observational period. Due to missing information on the volume of item ID 212020 I take an approximate value for typical fruit squashes of 0.2 L per bottle.

(e.g. employees or prices) varies over time, separately for the treatment (red line) and control group (blue dashed line).

My event studies (Figures 2, 4 and 6) analyse the difference in impacts of the tax between treatment and control group for each year separately. Put in other words, Figures 2, 4 and 6 show the differential change in employment and product prices of the treatment group compared to the control group. All of the subsequent event studies are set up uniformly. The x-axis reflects the years relative to the tax event, with zero being the policy year and for example, five the fifth year after the policy (i.e. 2023). I use 2018, the year of the policy introduction, as the reference year. The grey bars surrounding the coefficient estimates are confidence levels set at the 95 percent level.

4.1. Microeconomic Employment Analysis

Figure 2 shows the average logarithmic number of employees in soft drink companies (red line) and alcoholic beverage companies (blue dashed line), relative to the reference

year (i.e. 2017). The graph underlines that both groups followed similar trends in the pre-tax period. Only a short decrease in employment levels of alcohol producing firms can be observed from 2013 until 2014, compared to the employment levels in the reference year. Thereafter, both groups depict similar rising trends in employment until one year after the tax. Whilst employment of soft drink companies seems to continuously grow also hereafter, the employment level of alcohol producers fluctuates.

Figure 3 visualises the differential changes in logarithmic employment numbers²¹ of soft drink companies (treatment group) compared to alcoholic beverage companies (control group). The graph underlines that there are no significant differences in employment changes between soft drink companies compared to alcoholic beverage companies before the soft drink tax was implemented (see confidence intervals intersecting with the x-axis). Furthermore, overall changes in

²¹ Hereafter only referred to as employment.

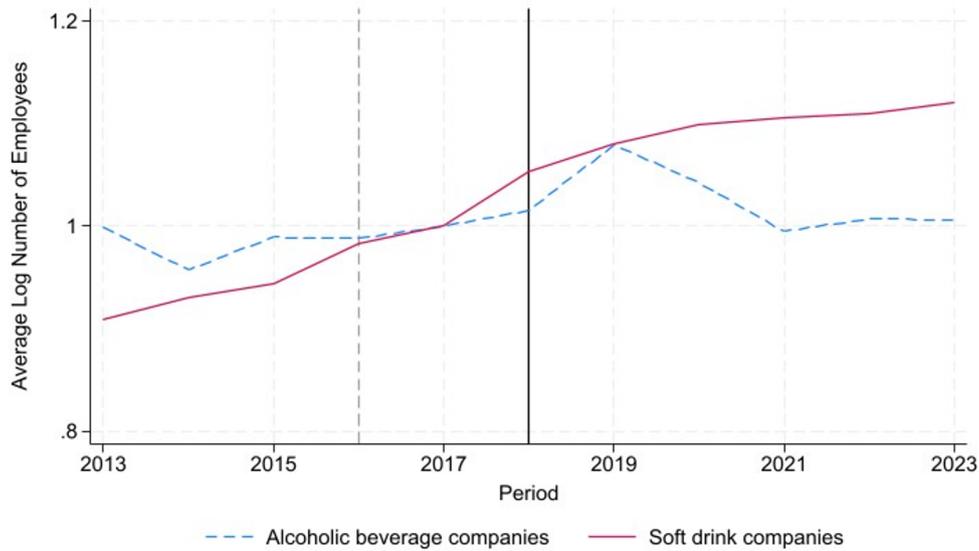


Figure 2: Parallel Trends - Microeconomic Employment

Note. Author’s own overview based on employment data retrieved from Refinitiv (2024b). Graph shows the average logarithmic number of employees in soft drink companies (treatment group) and in alcoholic beverage companies (control group) from 2013 to 2023. Logarithmic number of employees is normalised to one in 2017 (i.e. one year before the tax implementation).

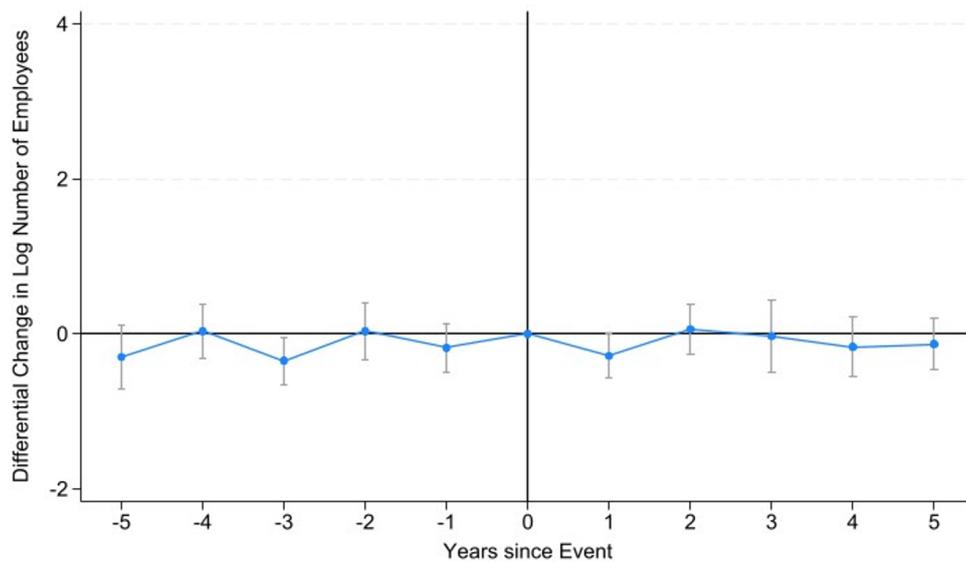


Figure 3: Event Study - Microeconomic Employment

Note. Author’s own overview based on employment data retrieved from Refinitiv (2024b). Graph shows the difference in logarithmic employment changes between soft drink companies (treatment group) and alcoholic beverage companies (control group). Confidence intervals are plotted at 95 percent level.

employment between soft drink firms and alcohol producing companies also do not seem significant in the period after the tax. Put differently, the event study suggests that the soft drink tax did not lead to any significant differences in employment changes between both groups in the post-tax period.

The results of my microeconomic employment regression analysis can be retrieved from Table 5. Column (1) reports the regression results without controlling for any

confounders. The regression estimates of my fully specified model (see Equation (1)) are depicted in Column (2). My coefficient of interest of the interaction term $Tax_t * Treatment_i$ (also known as difference-in-differences estimator) depicts the tax-induced difference in employment changes in soft drink companies and alcoholic beverage firms. The results in Column (1) imply that the soft drink tax leads to a significant positive change in employment in soft drink companies relative to employment changes in alcoholic beverage com-

Table 5: Employees Microeconomic Regression

Log Employees	(1)	(2)
Treatment*Tax	1.700*	0.089
	(0.860)	(0.062)
Treatment	0.069	
	(1.722)	
Tax	-1.021	-0.012
	(0.782)	(0.046)
Unemployment Rate		-0.076*
		(0.040)
GDP		-0.000
		(0.000)
Total Population		-0.000
		(0.000)
Lag Log Employees		0.871***
		(0.056)
Company FE	No	Yes
Observations	110	106
Adjusted R ²	0.063	0.995

Note. Author's own estimations. Column (1) depicts the coefficient estimates of regressing the tax event on logarithmic number of employees. The fully specified model in Column (2) additionally controls for *Unemployment Rate*, *GDP*, *Total Population* and company fixed effects. Standard errors are heteroskedasticity-robust and clustered at the company level. Standard errors in parentheses * p < 0.10, ** p < 0.05, *** p < 0.01.

panies. However, these results are negligible, due to the low explanatory power of the unadjusted model (e.g. no control variables). Although the difference-in-differences estimator in Column (2) also indicates a positive change in employment, these results are not significant. The control variable *Unemployment Rate* has a significant negative effect on the change in employment of soft drink firms, which is intuitive as higher national unemployment can lead to less employment in soft drink companies. *GDP* and *Total Population* have close to no effect at all. Overall, the UK soft drink tax did not have a significant impact on the number of employees in soft drink companies, when analysing microeconomic data. Hence, these results could be a first indication that the tax burden, which arises from the new soft drink taxation law, is not borne by employees in soft drink firms (e.g. in the form of staff reductions).

4.2. Macroeconomic Employment Analyses

4.2.1. Soft Drink Manufacturing Industry

To validate my above microeconomic employment results, I now use macroeconomic data, to again analyse whether the tax has an effect on employees in the soft drink manufacturing industry. Figure 4 depicts the average logarithmic number of employees²² in the soft drink manufactur-

ing industry (red line) and alcoholic beverage manufacturing industry (blue dashed line), relative to the reference year (2017). Again, I observe similar trajectories for both groups in the pre-tax period, strengthening my parallel trends assumption. Interestingly, in the year of the tax announcement (2016) I observe a small hike in employment levels for the soft drink manufacturing industry. This indicates that solely the announcement of the tax could have had an impact on employment. Anticipating the new tax burden for their high sugary drinks, soft drink companies might have taken precautionary measures before the tax even came into effect. For example, by hiring new employees for product innovation and reformulation tasks. This phenomenon is known as the Ashenfelter Dip, where the treatment group already alters their behaviour before the (policy) event occurs, as they expect to be undertaken a treatment soon (Lee, 2016). To control for this behaviour, I perform a robustness check using the announcement of the soft drink tax in 2016 as the event indicator (see Chapter 5). I find that my results remain robust. In general, Figure 4 indicates that both groups exhibit a continuously increasing trend in employment over the complete observational period.

The subsequent event study (Figure 5) further underlines the parallel trends assumption for the treatment and control group. All coefficients in the pre-tax period are insignificant

²² Hereafter only referred to as employment levels.

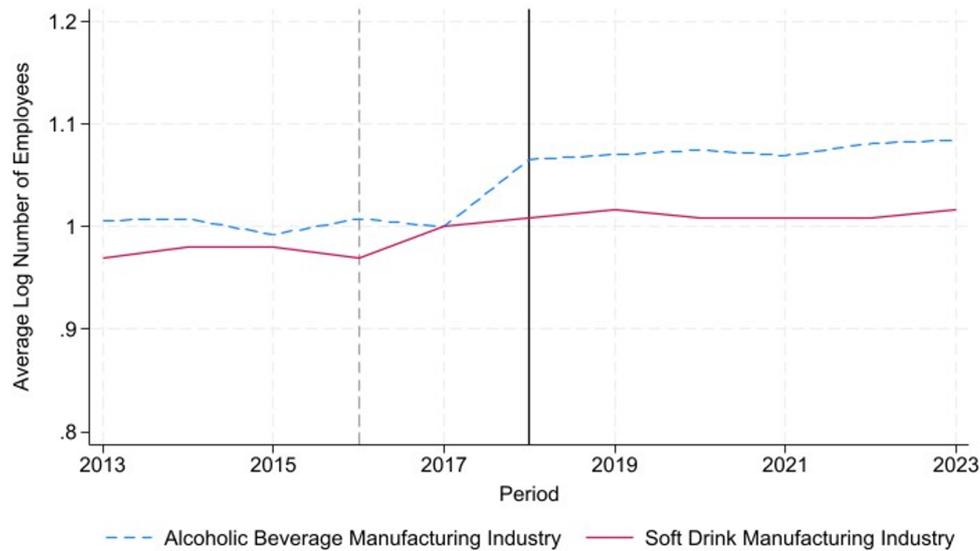


Figure 4: Parallel Trends - Macroeconomic Employment (Soft Drink Industry)

Note. Author’s own overview based on employment data retrieved from Office for National Statistics (2015, 2022b). Graph shows the average logarithmic number of employees in the soft drink manufacturing industry (treatment group) and in the alcoholic beverage manufacturing industry (control group) from 2013 to 2023. Logarithmic number of employees is normalised to one in 2017 (i.e. one year before the tax implementation).

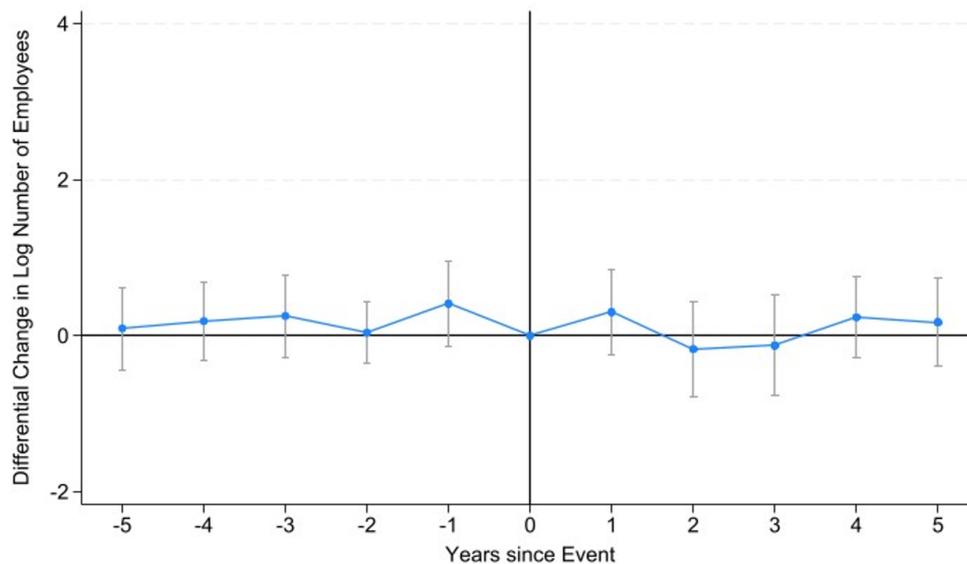


Figure 5: Event Study - Macroeconomic Employment (Soft Drink industry)

Note. Author’s own overview based on employment data retrieved from Office for National Statistics (2015, 2022b). Graph shows the difference in logarithmic employment changes between the soft drink manufacturing industry (treatment group) and alcoholic beverage manufacturing industry (control group). Confidence intervals are plotted at 95 percent level.

(see confidence intervals intersecting with the x-axis), indicating that there is no significant difference between employment changes in the soft drink manufacturing industry and alcoholic beverage manufacturing industry. This pattern pertains even after the soft drink tax was implemented, implying that the policy did not lead to significant changes in employment levels of the soft drink industry, relative to employment changes in the alcoholic beverage manufacturing industry.

4.2.2. Beverage Manufacturing Industry

Employment in the beverage manufacturing industry seems to exhibit the same pattern as in the preceding analysis of the soft drink industry. Figure 6 shows the average logarithmic number of employees²³ in the textile (control group) and beverage manufacturing industries (treatment group). Employment in the beverage manufacturing indus-

²³ Hereafter only referred to as employment.

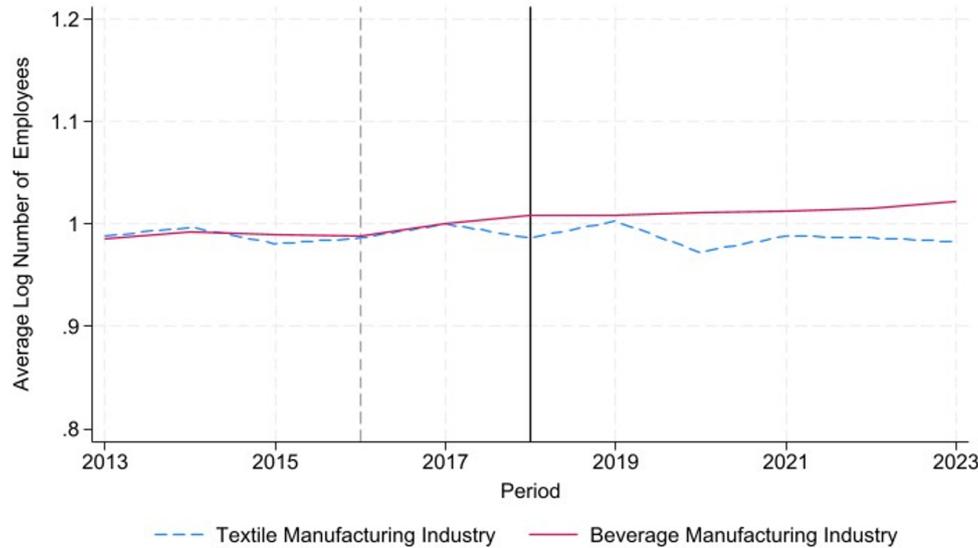


Figure 6: Parallel Trends - Macroeconomic Employment (Beverage Industry)

Note. Author's own overview based on employment data retrieved from Office for National Statistics (2015, 2022b). Graph shows the average logarithmic number of employees in the beverage manufacturing industry (treatment group) and in the textile manufacturing industry (control group) from 2013 to 2023. Logarithmic number of employees is normalised to one in 2017 (i.e. one year before the tax implementation).

try continuously rises over the entire observational period, compared to its value in the reference year (2017). Furthermore, the beverage (treatment) and textile industries (control) both have a similar positive pre-tax trend in employment, indicating strong parallel trends. Whilst the textile industry exhibits fluctuations over time, the beverage manufacturing industry demonstrates a continuous positive trend in employment, even after the tax implementation.

The event study of the beverage and textile manufacturing industries is depicted in Figure 7. Whilst the coefficient estimates seem slightly statistically significant in the first two years of the observational period (i.e. 2013 and 2014), they become insignificant afterwards. Put differently, the textile and beverage manufacturing industries have comparable logarithmic employment levels after 2015 onwards, making them valid treatment and control groups. Again, coefficient estimates for 2018 onwards are also statistically insignificant, suggesting that the soft drink tax did not lead to a significant change in employment levels of the beverage industry, compared to the change in employment levels in the textile industry.

The subsequent Table 6 presents the regression estimates for both analyses. First, I regress the tax event on the logarithmic number of employees in the soft drink manufacturing industry. Results are depicted in Columns (1) and (2). I then estimate the effect of the soft drink tax on the entire beverage manufacturing industry, of which results are presented in Columns (3) and (4). Columns (1) and (3) represent the coefficient estimates of regressing the tax on employment in the respective industries, excluding all control variables and fixed effects. The regression estimates of my adjusted models, as described in Equations (1) and (2), are

shown in Columns (2) and (4).

My regression coefficient of interest, the difference-in-differences estimator, is positive when regressing the soft drink tax on logarithmic employment in the soft drink manufacturing industry in Column (1). This implies a positive change in employment in the soft drink industry, compared to the employment change in the control group. When including controls, this induced change in employment in the soft drink industry becomes negative, relative to the control group. However, both effects are statistically insignificant. Thus, my macroeconomic regression results of employment in the soft drink industry are in line with my insignificant microeconomic regression result. Therefore, I can reject Hypothesis 1, which postulates that the UK soft drink tax leads to a decrease in employment in the soft drink manufacturing industry. Put differently, my results emphasise that employees in the soft drink manufacturing industry did not bear the cost of the soft drink tax, as employment levels have not significantly decreased due to the tax implementation. A possible explanation could be that either firms might have absorbed the extra costs of the new soft drink tax burden or have passed the costs onto consumers (see Section 4.3 for price regression results). Alternatively, soft drink companies might have even bypassed the tax by reformulating their high sugary drinks just in time, before the tax came into effect. My results are in line with current literature on soft drink taxation and employment in other countries, such as Peru, Portugal and the US (Díaz et al., 2023; Gonçalves et al., 2024; Lawman et al., 2019). They all found no significant effect of soft drink taxes on employment in the soft drink manufacturing industry in their respective countries of analysis.

Contrary, when regressing the soft drink tax on employ-

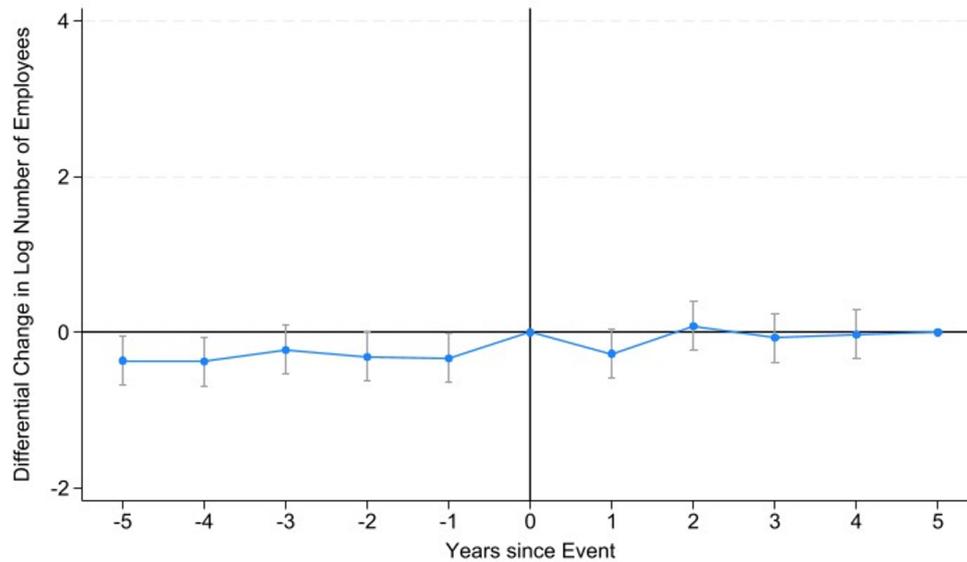


Figure 7: Event Study - Macroeconomic Employment (Beverage Industry)

Note. Author's own overview based on employment data retrieved from Office for National Statistics (2015, 2022b). Graph shows the difference in logarithmic employment changes between the beverage manufacturing industry (treatment group) and textile manufacturing industry (control group). Confidence intervals are plotted at 95 percent level. The last years' coefficient of the interaction term (Treatment * 2023) was omitted due to collinearity. Hence, only a coefficient estimate of zero is plotted without any confidence interval.

ment in the entire beverage manufacturing industry, results become highly significant (see Columns (3) and (4)). The difference-in-differences estimator of the basic unadjusted regression is positive and statistically significant (see Column (3)). Precisely, the coefficient of 0.274, which has to be back transformed first,²⁴ indicates that employment levels in the beverage manufacturing industry increased by 32% more than employment in the textile manufacturing industry. In Column (4) I see that this effect becomes even larger when controlling for confounding variables. Due to the UK soft drink tax, employment in the beverage manufacturing industry increased by 39%²⁵ more than employment in the textile industry. I also find that the *Unemployment Rate*, *GDP*, *Total Population* and the employment numbers of the previous year do not significantly influence employment levels, neither in the beverage nor in the soft drink manufacturing sector. Contrary to my initial assumption that employment levels will decline in response to the soft drink tax (see Hypothesis 1), I observe that employees in the beverage industry even benefitted from the tax, as employment has significantly increased in this sector (see Columns (3) and (4)).

To put these percentage changes into relation, absolute employment numbers before and after the tax implementation can be retrieved from Table 7. The absolute number of employees increased in both industries, the soft drink and the

overall beverage manufacturing industry, after the tax was implemented. However, when controlling for the general growing trends in the industries and other influential economic indicators (e.g. GDP) in my difference-in-differences analysis, only the employment increases in the beverage manufacturing industry are significant and attributable to the soft drink tax.

I ascribe this trend to non-soft-drink manufacturing companies in the beverage market, which might have benefitted from spillover effects. For example, alcoholic beverage manufacturers often diversify their product range, also offering non-alcoholic alternatives. By anticipating the new tax burden of their competitors (soft drink manufacturing firms), they might have created substitute products (e.g. low-sugar non-alcoholic beverages). Thus, these types of firms might have hired new employees for product innovation tasks and herewith increasing overall employment levels in the beverage manufacturing industry significantly.

Whilst other studies have found no significant employment effects (Díaz et al., 2023; Gonçalves et al., 2024; Lawman et al., 2019; Marinello, Leider, & Powell, 2021), I argue that the UK has a different setting. Compared to other countries with often uniform tax rates, the UK soft drink tax has a two-tiered tax design to encourage reformulation. I postulate that this tax policy increased overall employment levels in the beverage manufacturing industry, due to increased demand for employees to fulfil new product innovation tasks. Employment levels in the soft drink industry were not significantly affected. This indicates that they might have not required additional staffing for reformulation, as they might have solely passed on the tax to consumers. However, as ar-

²⁴ Coefficient estimates of log-linear models are interpreted as percentage changes (Wooldridge, 2013). In this case, the percentage change is calculated as follows: $(e^{0.274} - 1) * 100 = 31.52\%$.

²⁵ Again, for comprehension purposes, the percentage change is calculated as follows: $(e^{0.327} - 1) * 100 = 38.68\%$.

Table 6: Employees Macroeconomic Regression

Log(Employees)	Soft Drink Manufacturing Industry		Beverage Manufacturing Industry	
	(1)	(2)	(3)	(4)
Treatment*Tax	0.613 (0.529)	-0.129 (0.106)	0.274*** (0.068)	0.327*** (0.076)
Treatment	1.605*** (0.342)		-0.336*** (0.046)	
Tax	-0.320 (0.527)	0.360** (0.171)	-0.042 (0.058)	-0.089 (0.095)
Unemployment Rate		0.043 (0.096)		-0.044 (0.034)
GDP		-0.000 (0.000)		0.000 (0.000)
Total Population		0.000 (0.000)		-0.000 (0.000)
Lag Log Employees		0.354 (0.308)		-0.253 (0.237)
Observations	73	71	22	22
Industry FE	No	Yes	No	Yes
Adjusted R ²	0.082	0.957	0.703	0.706

Note. Author's own estimations. Coefficient estimates are presented separately for the dependent variable Log(Employees) in the soft drink manufacturing industry (Columns (1) and (2)) and in the beverage manufacturing industry (Columns (3) and (4)). Columns (1) and (3) depict the coefficient estimates of regressing the tax event on logarithmic number of employees. The fully specified (adjusted) models in Columns (2) and (4) additionally control for *Unemployment Rate, GDP, Total Population* and industry fixed effects. Standard errors are heteroskedasticity-robust but not clustered due to insufficient number of industry groups. Standard errors in parentheses * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 7: Average Absolute Number of Employees (Macroeconomic)

#	Soft Drink Manufacturing Industry	Alcoholic Beverage Manufacturing Industry	Beverage Manufacturing Industry	Textile Manufacturing Industry
Before Tax	10,000	5,151	37,400	52,400
After Tax	13,333	5,719	47,167	50,333
Absolute Difference	3,333	568	9,767	-2,067
Change in %	33.33	11.03	26.11	-3.94

Note. Author's own overview based on employment data retrieved from Office for National Statistics (2015, 2022b). The table shows the average number of employees in absolute values before (2013-2017) and after the tax was implemented (2018-2023) for each industry separately. The change in the average number of employees is depicted in percent.

gued above, I believe that alcoholic beverage manufacturers have utilised the soft drink tax introduction to expand their product portfolio. Thus, requiring additional employees and herewith increasing overall employment in the beverage sector.

4.3. Price Regression Analysis

As employees in the soft drink manufacturing industry did not have to bear the costs of the soft drink tax, and employees in the beverage manufacturing industry even benefitted from it, I now analyse whether the tax burden was passed onto the consumer. Thus, I regress the UK soft drink tax on soft drink prices. The following two graphs highlight the strength of my price regression analysis. Figure 8 shows the development of

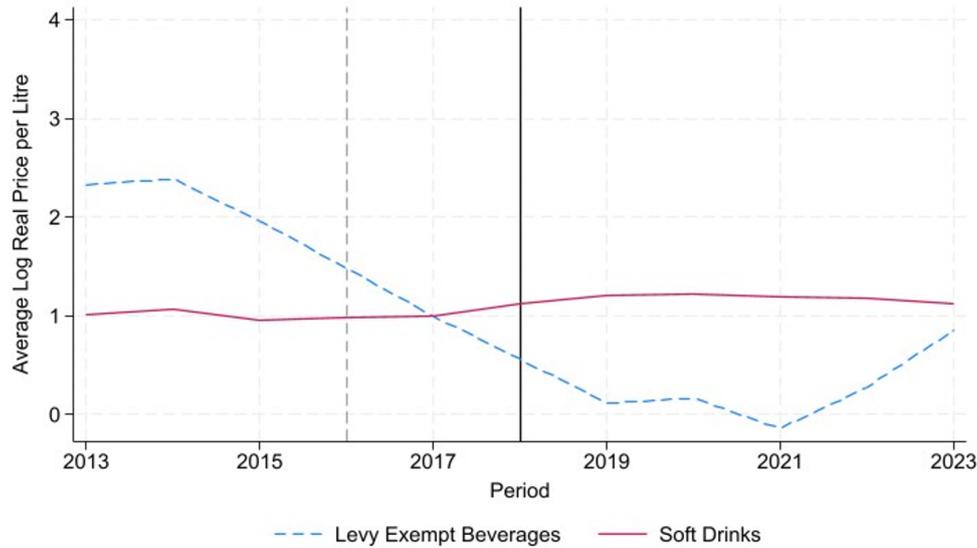


Figure 8: Parallel Trends - Prices

Note. Author's own overview based on price data retrieved from Office for National Statistics (2024b). Graph shows the average logarithmic real price per litre of soft drinks (treatment group) and levy-exempt beverages (control group) from 2013 to 2023. The logarithmic real price per litre is normalised to one in 2017 (i.e. one year before the tax implementation).

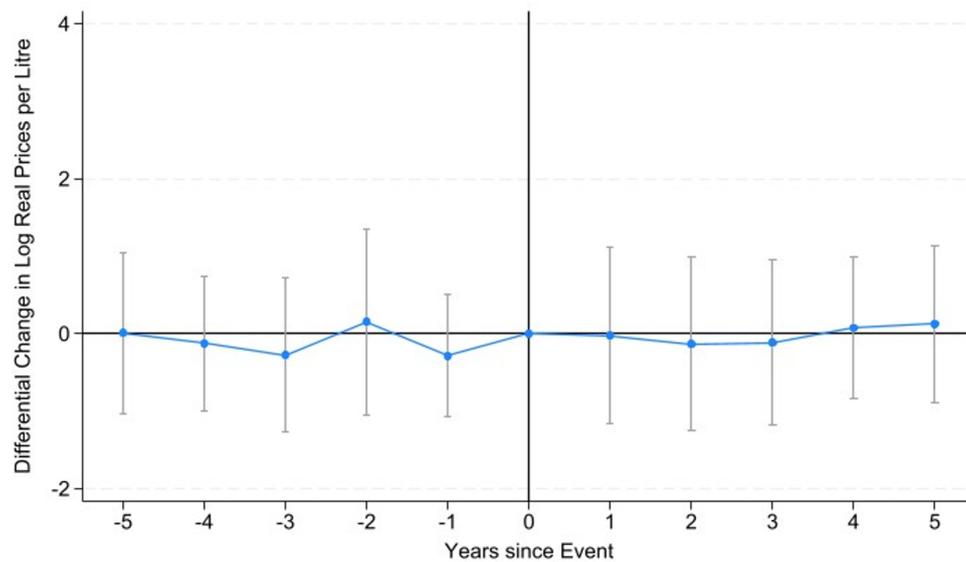


Figure 9: Event Study - Prices

Note. Author's own overview based on price data retrieved from Office for National Statistics (2024b). Graph shows the difference in logarithmic real price per litre changes between soft drinks (treatment group) and levy-exempt beverages (control group). Confidence intervals are plotted at 95 percent level.

the average logarithmic real price per litre²⁶ of soft drinks (red line) and levy-exempt beverages (e.g. juices and water) (blue dashed line), relative to the reference year 2017. The trend of soft drink prices seems to be relatively stable over the complete observational period and only showing a slight decline in the pre-tax period. However, the prices of levy-exempt beverages constantly decline until 2019 (and hereafter fluctuate) at a much higher rate.

I therefore conduct an event study to further investigate the parallel trends assumption. Figure 9 depicts the differential changes between the logarithmic real price per litre of levy-exempt beverages and soft drinks. As all coefficients are statistically insignificant, it can be assumed that there was no difference in price changes of levy-exempt beverages and soft drinks. This shows that the parallel trends assumption for treatment and control group holds. Overall, the event study also indicates that the soft drink tax probably does not have a statistically significant impact on the soft drink prices,

²⁶ Hereafter prices is used as synonym.

as the coefficients remain statistically insignificant after the tax came into effect.

As logarithmic percentage interpretations can be twofold at first sight in difference-in-differences analyses, I also provide an overview of average absolute prices before and after the tax implementation in Table 8. The overview shows that soft drink prices as well as prices of levy-exempt beverages in general decreased over my entire observational period between 2013 to 2023. Whilst the average soft drink price was £3.85 in the pre-tax period, soft drinks on average only costed £3.17 post tax, a total decrease of 68 pence. The same trend is also observable for levy-exempt beverages. To further test whether and to what extent this price change is attributable to the introduction of the soft drink tax, or whether these declines are a general trend of economic downturn, I conduct a difference-in-differences analysis.

Table 8: Average Absolute Prices

#	Soft Drink Prices	Levy-Exempt Beverages
Mean Price per Litre before tax (£ 2018)	3.85	1.34
Mean Price per Litre after tax (£ 2018)	3.17	1.09
Difference in Prices	-0.68	-0.25
Change in %	-17.7	-18.5

Note. Author's own overview based on price data retrieved from Office for National Statistics (2024b). The table shows the mean real price per litre in absolute values before (2013-2017) and after the tax (2018-2023) was implemented for treatment (soft drinks) and control group (levy-exempt beverages) separately. The change in average price is depicted in percent.

The results of my price regression analysis are reported in Table 9. I first regress the soft drink tax event solely on my dependent variable logarithmic real price per litre (see Column (1)). The difference-in-differences estimator indicates that the tax leads to a positive change in soft drink prices relative to the change in prices of levy exempt beverages. However, this effect is not statistically significant. When controlling for confounding variables, as well as for region and shop fixed effects, I estimate a statistically significant positive effect on price changes (see Column (2)). As I know that prices of soft drinks declined over the complete observational period, I revise the interpretation of the difference-in-differences estimator. As its coefficient depicts the difference in price changes between the treatment group compared to the control group, a positive coefficient either indicates that prices increased more or declined less in the treatment group than in the control group. In my case, I find that, despite the overall declining trend of beverage prices in the UK, the soft drink tax led to a 6%²⁷ less decline in prices of soft drinks

compared to prices of levy-exempt drinks. Put differently, even though the tax did not lead to significant price increases, it slowed down the general falling trend of soft drink prices. It can be said that, although the consumer did not visibly bear the soft drink tax burden in the form of higher product prices, I argue that they were still constrained by the tax, as soft drink prices would have declined even faster without the tax.

To quantify how much of the tax burden was passed onto the consumer, a pass through rate is estimated. However, as emphasised by Stacey et al. (2019), pass through rates can only be retrieved from regression analyses for uniform tax rates. For tiered tax rates, such as the UK soft drink tax, based on the sugar content and the applicable tax threshold, the pass-through rate could be calculated for each specific product. However, as my data does not provide nutritional information, I am not able to calculate the actual pass-through rate of the UK soft drink tax. However, I estimate the change in absolute prices, provided in Table 8, for comprehension purposes only. While Table 8 shows that soft drink prices (-17.7%) declined slightly less than levy exempt beverage prices (-18.5%), this does not depict the true causal effect of the soft drink tax. My difference-in-differences analysis, which controls for other influences on prices, finds that soft drink prices declined on average 6% less due to the soft drink tax and relative to levy-exempt beverage prices.

While declining prices seem counterintuitive at first, this trend could be caused by a general decline in demand for certain beverages, which might lead to significant promotions and herewith price decreases. Putting this into context, one of the few studies in this field, which also analyses the price effects of the UK soft drink tax, yields similar results. Using additional data on sugar content, Scarborough et al. (2020) differentiate between the two tax levies (i.e. the lower levy threshold has a tax rate of 18 pence per litre and the higher levy threshold of 24 pence per litre). They find that prices of soft drinks from the lower levy tier decreased by 10.7 pence, holding a tax pass-through rate of -59%. However, they also find that prices in the higher levy tier increased by 7.5 pence on average, equivalent to a pass-through rate of the tax onto consumers of 31%. They also identify an overall declining price trend for beverages in the control group, which costed 1.5 pence less 50 days post tax implementation.

Overall, my price results indicate that the UK soft drink tax decelerated the falling trend of soft drink prices. Whilst the general decrease in beverage prices is another research topic for itself, I argue that the beverage market is quite competitive, with a few large manufacturers predominating the market (see Figure 1). Thus, retail prices have to be kept as low as possible (e.g. by offering price promotions), in order for beverage companies to remain competitive. In view of the UK soft drink tax introduction, manufacturers were not able to hold soft drink prices as competitive as the prices of other levy-exempt beverages. This could be attributable to the new tax burden, which soft drink manufacturers partially passed onto consumers, in the form of fewer price promotions, compared to what they would have offered consumers, if the tax

²⁷ Again, for comprehension purposes, the percentage change is calculated as follows: $(e^{0.058} - 1) * 100 = 5.97\%$.

Table 9: Price Regression

Log(Price)	(1)	(2)
Treatment*Tax	0.036 (0.037)	0.058*** (0.022)
Treatment	0.700*** (0.099)	0.535*** (0.043)
Tax	-0.223*** (0.020)	-0.134*** (0.017)
Unemployment Rate		0.063*** (0.010)
GDP		0.000*** (0.000)
Total Population		0.000** (0.000)
Volume		-0.994*** (0.035)
Region FE	No	Yes
Shop Type FE	No	Yes
Observations	31,957	31,957
Adjusted R ²	0.142	0.587

Note. Author's own estimations. Column (1) depicts the coefficient estimates of regressing the tax event on logarithmic real price per litre (i.e. constant 2018 pounds). The fully specified model in Column (2) additionally controls for *Unemployment Rate*, *GDP*, *Total Population*, *Volume*, as well as region and shop type (i.e. chains with more than 10 shops or independent stores) fixed effects. Standard errors are heteroskedasticity-robust and clustered at shop code level (i.e. unique shop identifier, indicating from which shop the product prices were collected from). Standard errors in parentheses * p < 0.10, ** p < 0.05, *** p < 0.01.

would have not been implemented (see declining trend in soft drink prices before the tax).

An alternative explanation is that manufacturers did not have to pass the soft drink tax onto consumers, if they had reformulated their soft drinks early enough to bypass the levy. Scarborough et al. (2020) found that one year after the UK soft drink tax implementation, sugar in taxable soft drinks²⁸ was reduced by 2.13 grams per 100 millilitres on average. They observe that one year after the tax, 30.7% of the soft drinks that were tax liable (i.e. had a sugar content of 5 g per 100 ml or more), moved below the 5 g sugar threshold. This underlines that, in line with the UK soft drink tax objective, manufacturers substantially decreased the sugar levels in their products. Overall, I suggest that further analyses on the price effects associated with the UK soft drink tax need to be conducted to ensure more transparent results.

5. Robustness Checks

I test the validity of my price and employment regression results by performing a variety of adjustments to my main

²⁸ Referred to as intervention drinks in Scarborough et al. (2020).

model specifications (Equations (1) – (3)). Results can be derived from Tables A2 to A4 in the Annex. The tests are described in detail below. The first two checks are applied uniformly to all model specifications (Equations (1) - (3)). Robustness check three is applied uniformly to all employment analyses (Equations (1) - (2)). Hereafter, specific checks for price and employment analyses are conducted, adapting to the needs of the respective datasets.

1) Covid-19 Pandemic

The Covid-19 pandemic significantly impacted economies worldwide, and herewith not only affected employment but also prices due to high inflation. Thus, I re-run all regression analyses, limiting my observational period from 2013 to 2020, to account for the Covid crisis as confounding error (in employment and price estimates). All results are robust to this change (see Column (1) in Tables A2 - A4).

2) Tax Announcement Effect

As the UK soft drink tax was announced in 2016, two years before its actual implementation, soft drink manu-

facturers might have already adjusted their product prices or employment levels to compensate the anticipated future costs of the tax. This behaviour, also known as the Ashenfelter Dip, could distort regression results when the treatment group significantly changes their behaviour in the pre-tax period (Lee, 2016). Therefore, I use the tax announcement as regressor. I discover that only my macroeconomic soft drink analysis yields the same results as in the main specification. Results from my macroeconomic beverage industry analysis remain positive but become insignificant (see Column (7) in Table A3). Positive microeconomic employment estimates, which were insignificant before, become significant. This indicates that the announcement of the soft drink tax-induced increases in employment of soft drink companies, compared to alcoholic beverage companies. As this significant effect in the pre-tax period could distort my main microeconomic regression estimates (Table 5), results have to be treated with caution. My price estimates, which were significantly positive in the adjusted specification (Table 9), become insignificantly negative (see Column (2) in Table A4). Overall, this indicates that the announcement of the UK soft drink tax had no effect on soft drink prices.

3) Separate Estimations of Fixed Effects and Lagged Model Specifications

Angrist and Pischke (2009) emphasise that fixed effects regressions together with lagged dependent variables must be applied with caution due to their different assumptions. Thus, as suggested by the authors, I estimate each model separately. I apply this robustness check exclusively to my employment analyses, as no lagged variables are included in my price analysis due to the missing panel component of the data. The macroeconomic results remain robust. In Table A4, the microeconomic results remain positive when the lagged dependent variable is excluded (see Column (3)) but become insignificantly negative when excluding fixed effects (see Column (4)). However, I argue that the results of my main model specification (see Columns (1) and (2) in Table 5) are likely to be valid due to the subsequent reason. Angrist and Pischke (2009) postulate that separate estimations of fixed effects and lagged dependent models serve as a bracket in which the true causal effect lies. When the lagged dependent model specification is correct, but a fixed effects model is incorrectly applied, this leads to an overestimation of the actual effect (upper bound). Vice versa, if a lagged dependent model is incorrectly applied instead of the correct fixed effects model, this leads to an underestimation of the causal effect (lower bound) (Angrist & Pischke, 2009). Thus, my fixed effects estimation (see Column (3) in Table A2) is the upper bound and the lagged dependent variable model the lower bound (see Column (4) in Table A2), in which the true causal effect of my main model specification lies (see Column (2) in Table 5).

4) Employment Analyses 2016 to 2023

Furthermore, I control for the change in methodology of the Business Register and Employment Survey, from which my macroeconomic employment data is obtained. Whilst data between 2009 to 2015 only encompasses VAT registered businesses in Great Britain, data from 2016 to 2023 additionally includes smaller firms with less than 20 employees (i.e. Pay As You Earn [PAYE] businesses) (Office for National Statistics, 2017). To control for this inconsistency, I exclude all observations before 2016 in my macroeconomic employment analyses and find that my results are still robust (see Columns (5) and (10) in Table A3).

5) UK Primary Sales Market

My microeconomic employment analysis only includes businesses for which the UK is at least their third biggest sales market. I tighten this restriction in the robustness analysis to encompass only companies with the UK as primary sales market. I herewith exclude the largest multinationals, namely Coca-Cola Europacific Partners and Diageo. Again, my regression results are robust (see Column (5) in Table A2).

6) Missing Pre-Tax Data

In my microeconomic employment analysis, I control for the missing pre-tax data of companies. Hence, I exclude Artisanal Spirits, Adnams, Gusbourn and Distil from my analysis, and find that my initial results are valid (see Column (6) in Table A2).

For my price analysis I also exclude items with insufficient pre-tax observations. Thus, I exclude all cola drinks (item IDs 212011, 21025 and 21026), a bottle of mixer (item ID 212022) and a water bottle (item ID 21025). My initial results are also robust to this test (see Column (3) in Table A4).

7) "Pseudo" Panel Data

Lastly, to control for potential autocorrelation in my repeated cross-sectional price dataset, I construct an artificial ("pseudo") panel structure. I compute the average price of each item per period, resulting in 143 unique item-year observations. This robustness check yields similar results as my initial price regression (see Column (3) in Table A4).

6. Conclusion

This paper analyses the medium-term effects of the UK soft drink tax on employment and price levels. I find that neither consumers nor employees had to bear the cost of the soft drink tax visibly. However, despite the overall falling trend in beverage prices, consumers of soft drinks were minimally constrained by the tax, as they experienced a 6% less decline in soft drink prices (relative to levy-exempt beverages) than they would have without the tax implementation. Whilst the

tax had no significant impact on employment in the soft drink industry, I observe significant employment increases in the entire beverage industry. I postulate that alcoholic beverage manufacturers increased their staffing for product innovation purposes (e.g. low sugary non-alcoholic drinks), to attract soft drink consumers with their more competitive prices than those of their tax burdened competitors.

By using five years post-tax data, my study offers novel insights on the more manifested structural changes in the beverage sector, which has not been accounted in the numerous short-term studies before. I herewith provide evidence for policymakers and the UK government, showing that the soft drink tax did not come at the cost of employees and only minimally burdened consumers. Based on my results, the government can revise and adapt the policy to further incentivise reformulation and herewith unburden the consumer.

My study is limited in the following ways. Due to missing nutritional information, I was not able to distinguish the effect of the soft drink tax on lower levy and higher levy tier drinks. Hence, I also could not quantify the pass-through rate of the tax. Thus, I was also unable to identify as to how much firms are burdened with the tax or bypass it by reducing sugar in soft drinks just below the threshold (i.e. bunching). Another limitation is that I could not account for autocorrelation in prices, hence price estimates have to be interpreted with caution. Lastly, due to the lack of an adequate geographical control group, I was not able to control for potential spillover effects of the tax (e.g. shifts of employment from the soft drink industry into alcoholic beverage manufacturing industry).

As my study and the few other studies in this field diverge largely in estimations, I encourage further price analyses to arrive at a common understanding on the tax incidence of consumers. Also, further research on the general pricing structure in the UK beverage market needs to be conducted, to comprehend the unusual trend of the overall decline in beverage prices.

References

- Aguilar, A., Gutierrez, E., & Seira, E. (2021). The effectiveness of sin food taxes: Evidence from Mexico. *Journal of Health Economics*, 77, 102455. <https://doi.org/10.1016/j.jhealeco.2021.102455>
- Allcott, H., Lockwood, B. B., & Taubinsky, D. (2019). Regressive Sin Taxes, with an Application to the Optimal Soda Tax. *The Quarterly Journal of Economics*, 134(3), 1557–1626. <https://doi.org/10.1093/qje/qjz017>
- Alsukait, R., Wilde, P., Bleich, S., & Singh, G. (2020). Evaluating Saudi Arabia's 50% carbonated drink excise tax: Changes in prices and volume sales. *Economics & Human Biology*, 38(100868). <https://doi.org/10.1016/j.ehb.2020.100868>
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. <https://doi.org/10.2307/j.ctvc4j72>
- Bank of England. (n.d.). Daily spot exchange rates against US Dollars. Retrieved October 20, 2024, from <https://www.bankofengland.co.uk/boeapps/database/Rates.asp?Travel=NixAZx&into=USD>
- Briggs, A. D. M., Mytton, O. T., Kehlbacher, A., Tiffin, R., Elhussein, A., Rayner, M., Jebb, S. A., Blakely, T., & Scarborough, P. (2017). Health Impact Assessment of the UK Soft Drinks Industry Levy: A Comparative Risk Assessment Modelling Study. *Lancet Public Health*, 2(1), e15–e22. [https://doi.org/10.1016/S2468-2667\(16\)30037-8](https://doi.org/10.1016/S2468-2667(16)30037-8)
- Cameron, A. C., & Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2), 317–372. <https://doi.org/10.3368/jhr.50.2.317>
- Cawley, J., & Frisvold, D. (2017). The Pass-Through of Taxes on Sugar-Sweetened Beverages to Retail Prices: The Case of Berkeley, California: The Pass-Through of Taxes on Sugar-Sweetened Beverages. *Journal of Policy Analysis and Management*, 36(2), 303–326. <https://doi.org/10.1002/pam.21960>
- Cawley, J., & Frisvold, D. (2023). Review: Taxes on sugar-sweetened beverages: Political economy, and effects on prices, purchases, and consumption. *Food Policy*, 117(102441). <https://doi.org/10.1016/j.foodpol.2023.102441>
- Cuadrado, C., Dunstan, J., Silva-Illanes, N., Mirelman, A., Nakamura, R., & Suhrcke, M. (2020). Effects of a sugar-sweetened beverage tax on prices and affordability of soft drinks in Chile: A time series analysis. *Social Science & Medicine*, 245, 112708. <https://doi.org/10.1016/j.socscimed.2019.112708>
- Das, P. (2019). *Econometrics in Theory and Practice*. Springer Singapore. <https://doi.org/10.1007/978-981-32-9019-8>
- de la Feria, R. (2024). Non-(Fully) Harmonised Excise Taxes and Irrebuttable Presumptions. *EC Tax Review*, 33(3), 98–108. <https://ssrn.com/abstract=4775842>
- Department of Health and Social Care. (2022). New calorie labelling rules come into force to improve nation's health. <https://www.gov.uk/government/news/new-calorie-labelling-rules-come-into-force-to-improve-nations-health>
- Díaz, J.-J., Sánchez, A., Diez-Canessco, F., Jaime Miranda, J., & Popkin, B. M. (2023). Employment and wage effects of sugar-sweetened beverage taxes and front-of-package warning label regulations on the food and beverage industry: Evidence from Peru. *Food Policy*, 115, 102412. <https://doi.org/10.1016/j.foodpol.2023.102412>
- Doran, J., Bourke, J., & Kirby, A. (2017). *Introduction to Applied Econometrics Analysis Using Stata*. Oak Tree Press. <https://de.everand.com/book/359276951/Introduction-to-Applied-Econometrics-Analysis-Using-Stata>
- Gonçalves, J., Merenda, R., & Pereira dos Santos, J. (2024). Not so sweet: impacts of a soda tax on producers. *International Tax and Public Finance*, 31, 1388–1412. <https://doi.org/10.1007/s10797-023-09808-7>
- Guerrero-López, C., Molina, M., & Colchero, M. A. (2017). Employment changes associated with the introduction of taxes on sugar-sweetened beverages and nonessential energy-dense food in Mexico. *Preventive Medicine*, 105, 43–49. <https://doi.org/10.1016/j.ypmed.2017.09.001>
- Hattersley, L., Fuchs, A., Gonima, A., Silver, L., & Mandeville, K. (2020). *Business, Employment, and Productivity Impacts of Sugar-Sweetened Beverages Taxes* (tech. rep.). World Bank. <https://hdl.handle.net/10986/34082>
- Heiden, S. (2002). *Kursreaktionen auf Dividendenankündigungen: Ereignisstudie am deutschen Kapitalmarkt*. Deutscher Universitätsverlag. <https://doi.org/10.1007/978-3-322-89777-0>
- HM Revenue & Customs. (n.d.). Income Tax: introduction. Retrieved October 20, 2024, from <https://www.gov.uk/income-tax/how-you-pay-income-tax>
- HM Revenue & Customs. (2016). Soft Drinks Industry Levy. Retrieved July 6, 2024, from <https://www.gov.uk/government/publications/soft-drinks-industry-levy/soft-drinks-industry-levy>
- HM Revenue & Customs. (2024a). Soft Drinks Industry Levy statistics background and references. Retrieved November 6, 2024, from <https://www.gov.uk/government/statistics/soft-drinks-industry-levy-statistics/soft-drinks-industry-levy-statistics-background-and-references>
- HM Revenue & Customs. (2024b). Soft Drinks Industry Levy statistics: Commentary 2024. Retrieved November 6, 2024, from <https://www>

- .gov.uk/government/statistics/soft-drinks-industry-levy-statistics/soft-drinks-industry-levy-statistics-commentary-2021
- Imbens, G. W., & Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1), 5–86. <https://doi.org/10.1257/jel.47.1.5>
- Kumaranyake, L. (2000). The real and the nominal? Making inflationary adjustments to cost and other economic data. *Health and Policy Planning*, 15(2), 230–234. <https://doi.org/10.1093/heapol/15.2.230>
- Lauber, K., Rippin, H., Wickramasinghe, K., & Gilmore, A. B. (2022). Corporate political activity in the context of sugar-sweetened beverage tax policy in the WHO European Region. *European Journal of Public Health*, 32(5), 786–793. <https://doi.org/10.1093/eurpub/ckac117>
- Lawman, H. G., Bleich, S. N., Yan, J., LeVasseur, M. T., Mitra, N., & Roberto, C. A. (2019). Unemployment claims in Philadelphia one year after implementation of the sweetened beverage tax. *PLoS ONE*, 14(3), e0213218. <https://doi.org/10.1371/journal.pone.0213218>
- Lee, M. (2016). *Matching, Regression Discontinuity, Difference in Differences, and Beyond*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190258733.003.0005>
- London Stock Exchange. (2023). Audited full-year results. Retrieved October 2, 2024, from <https://www.londonstockexchange.com/news-article/EISB/audited-full-year-results/15938249>
- Marcus, M., & Sant'Anna, P. H. C. (2021). The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics. *Journal of the Association of Environmental and Resource Economists*, 8(2), 235–275. <https://doi.org/10.1086/711509>
- Marinello, S., Leider, J., & Powell, L. M. (2021). Employment impacts of the San Francisco sugar-sweetened beverage tax 2 years after implementation. *PLoS ONE*, 16(6), e0252094. <https://doi.org/10.1371/journal.pone.0252094>
- Marinello, S., Leider, J., Pugach, O., & Powell, L. M. (2021). The impact of the Philadelphia beverage tax on employment: A synthetic control analysis. *Economics & Human Biology*, 40, 100939. <https://doi.org/10.1016/j.ehb.2020.100939>
- Metcalf, S., & Sasse, T. (2023). *Tackling obesity* (tech. rep.). Institute for Government. <https://www.instituteforgovernment.org.uk/publication/tackling-obesity>
- NISRA. (2023). BRES Publication and Tables 2022. BRES 2022 Detailed Tables – Excel Format. Retrieved October 28, 2024, from <https://www.nisra.gov.uk/publications/bres-publication-and-tables-2022>
- Office for National Statistics. (2009). UK Standard Industrial Classification of Economic Activities 2007 (SIC 2007): Structure and explanatory notes. <https://www.ons.gov.uk/methodology/classificationstandards/ukstandardindustrialclassificationofeconomicactivities/uksic2007>
- Office for National Statistics. (2015). Business Register and Employment Survey (excluding units registered for PAYE only) : open access (2009-2015). Retrieved June 25, 2024, from <https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=172>
- Office for National Statistics. (2016). Consumer Price Inflation Basket of Goods and Services: 2016. <https://www.ons.gov.uk/economy/inflationandpriceindices/articles/consumerpriceinflationbasketofgoodsandservices/2016>
- Office for National Statistics. (2017). Business Register and Employment Survey (BRES). Retrieved August 24, 2024, from <https://www.nomisweb.co.uk/articles/1062.aspx>
- Office for National Statistics. (2018). CPIH Microdata Metadata Template. Retrieved November 10, 2024, from <https://www.ons.gov.uk/economy/inflationandpriceindices/adhocs/008009cpihmicrodatametadata-template>
- Office for National Statistics. (2020). A guide to labour market statistics. <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/aguidetoemploymentandlabourmarketstatistics>
- Office for National Statistics. (2022a). Business Register and Employment Survey. Retrieved October 29, 2024, from <https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/businessregisterandemploymentsurvey>
- Office for National Statistics. (2022b). Business Register and Employment Survey : open access (2015 to 2022). Retrieved June 25, 2024, from <https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=189>
- Office for National Statistics. (2023). Consumer Price Inflation (includes all 3 indices – CPIH, CPI and RPI) QMI. <https://www.ons.gov.uk/economy/inflationandpriceindices/methodologies/consumerpriceinflationincludesall3indicescpihcpiandrpiqmi>
- Office for National Statistics. (2024a). Consumer Price Inflation Basket of Goods and Services: 2019 - 2021 edition of this dataset Tables and annexes. <https://www.ons.gov.uk/economy/inflationandpriceindices/datasets/consumerpriceinflationbasketofgoodsandservices>
- Office for National Statistics. (2024b). Consumer price inflation item indices and price quotes. Retrieved June 26, 2024, from <https://www.ons.gov.uk/economy/inflationandpriceindices/datasets/consumerpriceindicescpiandretailpricesindexrpiitemindicesandpricequotes>
- Office for National Statistics. (2024c). Unemployment rate (aged 16 and over, seasonally adjusted): %. Retrieved July 20, 2024, from <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/unemployment/timeseries/mgsx/lms>
- Refinitiv. (2023). *Refinitiv Global Equity Indices* (tech. rep.). London Stock Exchange Group. https://www.lseg.com/content/dam/marketing/en_us/documents/methodology/global-equity-index-methodology.pdf
- Refinitiv. (2024a). *Constituents and Market Capitalisation of the UK Beverage Sector Index* [Retrieved 12 June, 2024, from Refinitiv Eikon Terminal.].
- Refinitiv. (2024b). *Number of Employees Variable WC07011, 2013-2023* [Retrieved 12 June, 2024, from Refinitiv Eikon Terminal.].
- Reuters. (n.d.). Momentous Holdings Corp (MMNT.PK). Retrieved October 26, 2024, from <https://www.reuters.com/markets/companies/MMNT.PK/>
- Scarborough, P., Adhikari, V., Harrington, R. A., Elhoussein, A., Briggs, A., Rayner, M., Adams, J., Cummins, S., Penney, T., & White, M. (2020). Impact of the announcement and implementation of the UK Soft Drinks Industry Levy on sugar content, price, product size and number of available soft drinks in the UK, 2015-19: A controlled interrupted time series analysis. *PLoS Medicine*, 17(2), e1003025. <https://doi.org/10.1371/journal.pmed.1003025>
- Stacey, N., Mudara, C., Wen Ng, S., van Walbeek, C., Hofman, K., & Edoka, I. (2019). Sugar-based beverage taxes and beverage prices: Evidence from South Africa's Health Promotion Levy. *Social Science & Medicine*, 238, 112465. <https://doi.org/10.1016/j.socscimed.2019.112465>
- Statista. (2024). Leading beverage companies by market capitalization in the United Kingdom (UK) as of October 15, 2024. Retrieved October 31, 2024, from <https://www.statista.com/statistics/1345274/uk-largest-beverage-companies-by-market-capitalization/>
- United Nations. (2018). *Classification of Individual Consumption According to Purpose (COICOP) 2018* (Statistical Papers Series M No. 99). https://unstats.un.org/unsd/classifications/unsdclassifications/COICOP_2018_-_pre-edited_white_cover_version_-_2018-12-26.pdf
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th). South-Western, Cengage Learning.
- Wooldridge, J. M. (2021). *Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators* (Working Paper). <https://doi.org/10.2139/ssrn.3906345>
- World Bank. (2020). *Taxes on Sugar-Sweetened Beverages: International Evidence and Experiences* (tech. rep.). <https://openknowledge.worldbank.org/entities/publication/9498c3ef-b425-54f1-aafd-d3d3b8411674>
- World Bank. (2023a). GDP per capita (current US\$) - United Kingdom. Retrieved July 25, 2024, from <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=GB>

- World Bank. (2023b). Population, total - United Kingdom. Retrieved July 25, 2024, from <https://data.worldbank.org/indicator/SP.POP.TOTL?locations=GB>
- World Health Organization. (2023a). *Global report on the use of sugar-sweetened beverage taxes 2023* (tech. rep.). <https://www.who.int/publications/i/item/9789240084995>
- World Health Organization. (2023b). Noncommunicable diseases. Retrieved July 20, 2024, from <https://www.who.int/en/news-room/fact-sheets/detail/noncommunicable-diseases>
- Young, S. (2016). Fever-Tree Aims to Mix It With Drinks Industry Elite. *Reuters*. Retrieved October 25, 2024, from <https://www.reuters.com/article/world/uk/fever-tree-aims-to-mix-it-with-drinks-industry-elite-idUSKCN0Y30WQ/>



From Pictures to Perceptions: Exploring the Strategic Use of Visuals in CSR Reports and the Impact of Regulatory Mandates

Thiemo Janßen

Technical University of Munich

Abstract

This thesis explores the strategic use of visuals in CSR reports, comparing companies in the EU and Switzerland. Using automated image classification and clustering, 11,455 images from sustainability reports were analyzed. The study finds that firms in environmentally or socially sensitive industries, especially those with higher CO₂ emissions per revenue, tend to use more images—often aligned with their specific challenges. This suggests that visuals are not merely illustrative but serve to shape corporate perception, potentially diverting attention from negative impacts. The introduction of the EU's Non-Financial Reporting Directive (NFRD) appears to reduce image reliance, indicating a positive regulatory effect on transparency. Overall, the findings highlight that images can be used not just to complement text, but to construct a more favorable corporate narrative. The study underscores the importance of critically assessing visual elements in CSR disclosures, as they may subtly influence stakeholders' perception beyond what is verbally communicated.

Keywords: CSR; image analysis; NFRD; greenwashing; visual communication

1. Introduction

One of the most dominant topics in today's world, both for institutions and for society in general, is the advance of global warming and its consequences (Carleton & Hsiang, 2016). Looking at the most significant contributors to this development, we find that companies and their actions emit a vast majority of the responsible global greenhouse gas emissions. The industry sector alone is responsible for 37 percent of these emissions (Worrell et al., 2009). This reality has led to a heightened focus on the environmental activities of companies and, in the shadow of this development, also on the responsible behavior of companies in general, often referred to as CSR (I use the term CSR interchangeably with ESG throughout this paper. When I mention sustainability, I also mean the realm of CSR¹). This led to stakeholders calling

for greater transparency and accountability through comprehensive reporting (Huang & Kung, 2010)

Due to this increasing pressure on companies, we have seen a sharp increase in the number of companies issuing sustainability reports (Serafeim & Amel-Zadeh, 2017; Stolowy & Paugam, 2018). The qualitative nature of CSR (compared to the more quantitative nature of financial reports) and the fact that different industries are affected differently pose challenges to the readers in extracting the meaningfulness of these reports and comparing them to other companies (Christensen et al., 2021; Serafeim & Amel-Zadeh, 2017). Some countries respond even further to this call for transparency by introducing reporting mandates, such as the EU's Non-Financial Reporting Directive (NFRD). The goal is to increase the comparability and transparency of these reports (Directive 2014/95, recital 1) and create an incentive to drive sustainability efforts (Directive 2014/95, recital 3). While re-

¹ It is important to note, though, that CSR generally refers to a company's voluntary actions to contribute positively to society and the environment, emphasizing ethical behavior and social impact. ESG, on the other hand, is a framework for assessing a company's impact in specific areas of en-

vironment, social involvement, and governance, often linked with how these are integrated into the company's operational and strategic management."

porting can increase transparency, companies also use reports for their own interest, which can manipulate readers' perceptions (Siano et al., 2017). A theory applied to explain this is Legitimacy Theory, which argues that companies use reports to justify their actions and convey that they act according to the greater good of society (Boiral & Henri, 2017; Cho et al., 2009; Suchman, 1995). Following this, it can be reasoned that the use of, e.g., boilerplate language or the deliberate framing of reports can be used to create a positive image of the company while it does not change its behavior (Boiral, 2013; Christensen et al., 2021). The qualitative nature of CSR reports mentioned before makes falsifying claims and distinguishing between actual information even harder. While it could be argued that a mandate would solve this problem by establishing standards, it has been shown in the past that reporting mandates often increase the use of boilerplate language (Dyer et al., 2017).

Besides boilerplate language, the use of images in CSR reports is particularly interesting in this context. Visuals can be potent tools for influencing perception, as they often shift the focus towards them (Tversky, 1974), subconsciously affect one's perception (Posner et al., 1976), and do not offer quantifiable data that can be critically assessed.

This raises the question of how and why companies use images in their CSR reports. To address this, I first investigate the implementation of CSR reporting, its historical context, and the implications of introducing mandates like the NFRD. This is followed by an examination of trends in CSR reporting behaviors in general and the role of visual elements, together with their influence on stakeholder perception.

Building on this foundation, the thesis will explore which specific characteristics of companies influence their use of images in CSR reports and whether a difference emerges in the presence of reporting mandates. By comparing the European Union, with its NFRD mandate, and Switzerland, which lacks such a mandate, we will gain insights into the impact of mandates on the use of visuals within these reports. The reason behind comparing the EU with Switzerland is that those countries, apart from the mandate, face similar societal pressures and operate in a similar environment.

For stakeholders to effectively evaluate CSR Reports, it is essential to have a comprehensive understanding of the underlying motivations and contexts that drive the use of visuals within these documents. Images possess the ability to influence viewers, often on a subconscious level. They can either highlight achievements or divert attention away from unfavorable elements. It is essential to be able to identify instances where images are employed not for clarity but rather to manipulate perception.

2. CSR Reporting: Why it Matters, History and the NFRD

2.1. Increasing focus on CSR

In the corporate landscape, stakeholders play a crucial role in shaping the activities of companies, including those related to CSR. Companies are increasingly expected to operate

sustainably and ethically. An example of this development is the Business Roundtable's 2019 statement on the purpose of business. Top US executives decided that a company's primary purpose is not only to serve its shareholders but also its customers, employees, suppliers, and communities (Business Roundtable, 2019). "These executives are responding to mounting pressure that a company needs to do 'good' while doing business, whether that means keeping carbon emissions low, waterways clear, or workers healthy" (Christensen et al., 2021, p. 1177). From a more general perspective, the 17 Sustainable Development Goals (SDGs), a core part of the United Nations (UN) Agenda 30, also indicate a shift in what is expected from companies by governments and society, detailing the topics a CSR initiative can or should be centered around. Even though these are not explicitly published to guideline cooperation's they suggest an attitude change and an enhanced awareness of sustainability topics from society and consequently the consumers.

That consumers are an integral part of a company's business is self-evident. However, in the realm of CSR, consumers also play the most significant role in shaping corporate actions towards CSR activities. A study from Accenture revealed that 68% of CEOs acknowledge consumers as the most influential stakeholder group concerning their sustainability agenda (followed by governments and investors) (Accenture, 2022).

The investor's focus on CSR can be seen in a study conducted by BlackRock (2020). They ascertain that there has been an increase in Assets under Management related to CSR-conforming products from 895 billion in 2015 to 1833 billion in 2020 (only up to September). They further state that for 78% of all investors surveyed, CSR is a central component of their strategy. While the most prominent reason Investors adopt sustainability investing is that it is the "right thing to do", the second and third most mentioned reasons are risk-based (BlackRock, 2020). From this, you can argue that investors recognize the pressure from society and consumers on companies, as misbehavior could negatively affect a company's business.

2.2. CSR reporting as a communications tool

To mitigate these pressures, companies use CSR reporting to convey legitimacy to their stakeholders (Cho et al., 2009). Legitimacy in this sense is defined as "a generalized perception or assumption that an entities actions are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions" (Suchman, 1995, p. 574), and thus serving as a critical pillar in corporate communication strategies and the maintenance of a positive corporate image. This concept can also be observed in traditional financial reporting. Davison & Warren highlight that "Annual reports are almost universally used as a means of molding corporate identity and reputation" (Davison & Warren, 2009, p. 846).

Furthermore, Erickson et al. (2011) examine how companies use reporting to strategically manage information, especially during times of crisis or when facing adverse events.

They found that reports can be used to shape stakeholder perception and reduce potential adverse reactions, thereby preserving legitimacy and a positive corporate image. Voluntary reporting can also serve as a means to decrease external costs or pressure from stakeholders, allowing companies to proactively defend and restore their image in the case of misbehavior or wrongdoing (Tate et al., 2009). In essence, CSR reporting is a strategic tool to enhance legitimacy and corporate image by demonstrating its commitment to sustainability, addressing expectations, and potentially altering perceptions of its performance and ethical stance (Boiral & Henri, 2017). From this perspective, it can be explained why companies that don't perform well still find it essential to report on CSR. By reporting, they seek to maintain trust and credibility with their stakeholders (Nielsen & Thomsen, 2018).

As mentioned in the preceding chapter, investors' growing focus on CSR can be seen as another significant factor for companies to issue CSR reports. As more capital is invested in ESG-conforming assets, these investors consequently demand information about this realm. Not being transparent by not disclosing CSR information would lead investors to neglect the company when searching for ESG-conforming investments. Hence, CSR reporting also becomes more critical from a classical shareholder perspective.

Therefore, the increase in public awareness and pressure is also reflected in the number of companies that publish CSR reports seeking to legitimize their activities. The share of companies disclosing CSR information rose between 2002 and 2015 from 10% to 80% in the EU (regarding companies from the EuroStoxx600) and from 10% to 60% in the US (regarding companies in the S&P 500) (Stolowy & Paugam, 2018). Another study found that between the early 1990s and 2016, the number of companies issuing some form of sustainability reporting increased from around 20 to almost 900 (Serafeim & Amel-Zadeh, 2017).

2.3. Differences from traditional reporting

Despite their similarity in function, traditional (financial) reporting and CSR reporting differ significantly in many aspects. In the following section, we will look at some key differences.

First and most apparent, externalities play a central role in CSR reporting. The information disclosed is not only about the company itself but also its environmental and societal impact. It could also extend outside traditional boundaries, for example, "when a firm imposes child labor restrictions on its supply chain" (Christensen et al., 2021, p. 1186).

Furthermore, the readership of CSR reports is potentially much larger than that of financial reporting. As discussed in the previous chapter, there are more stakeholders engaged in the CSR dimension of a company, making it more challenging to address all the interests of these diverse stakeholders. Moreover, some stakeholders, e.g., customers, are less experienced in reading corporate reports and use the information for different purposes than traditional investors, mainly evaluating the company's contribution and adherence to their norms and values (Christensen et al., 2021).

The diversity of topics addressed by the reports is another differentiating factor. As the concepts of CSR and sustainability are not clearly defined, they include a wide range of ESG topics, actions, and policies. They also differ substantially between different industries, nations, and companies (Christensen et al., 2021). Each company might implement different solutions or approaches to address CSR-related topics. While many of these activities might result in technical or measurable outputs (like the total amount of CO2 Emissions), they cannot be measured in traditional monetary terms. Also, there are still other actions or activities that cannot be quantified at all. CSR is often seen as a strategic endeavor that prioritizes long-term advantages for a company over immediate financial gains (Benabou & Tirole, 2010); these long-term activities are often qualitative (Christensen et al., 2021). This situation makes it difficult to apply standard reporting practices like double-entry bookkeeping, resulting in a variety of different formats, making comparisons and standardization difficult (Christensen et al., 2021; Kitzmueller & Shimshack, 2012).

Due to the diversity of users and topics, a broad spectrum of interests and preferences originates internally and externally from the company. As a result, CSR reports serve a variety of functions that might differ over time, for example, when social activists target the firm or when exogenous shocks, such as a natural catastrophe, occur (Baron, 2001; Bonetti et al., 2023; Christensen et al., 2021). When targeted by a group (or when conceived to be targeted), the company might strategically use the report to (proactively) defend itself against (potential) claims.

Stemming from these differences, the key issues are the heterogeneity of the published reports and the hardly quantifiable metrics. This heterogeneity leads to difficulties in comparing companies against each other (Christensen et al., 2021), which is one of the central goals that readers, especially investors, try to accomplish when evaluating CSR reports. This is highlighted by a BlackRock study where Investors stated that the biggest challenge for them to adopt sustainable investing is the "Poor quality or availability of ESG data and analytics" (53% of respondents) and "Poor quality of sustainability investment reporting" (33% of respondents) (BlackRock, 2020). Serafeim and Amel-Zadeh (2017) also support this narrative; they point out that the biggest challenge to using ESG information is the lack of comparability. Further, they also suggest that the quantification and reliability of the reported information are major challenges.

Several developments have tried to standardize reporting practices and establish greater comparability to address this issue, mainly through the development of guidelines and reporting mandates. These are expected to improve homogeneity, at least within an industry (Christensen et al., 2021).

2.4. Homogenization approaches

As the need for detailed and homogenous CSR disclosures increases, several organizations have initiated efforts to

establish voluntary reporting standards to enhance the consistency of reports published. For example, the Sustainability Accounting Standards Board (SASB) developed “industry-specific disclosure standards across financially material environmental, social, and governance topics” (Christensen et al., 2021, p. 1177). This approach focuses more on the integration of ESG reporting into financial reporting. This standard aims to guide companies in 77 different industries, allowing them to identify, manage, and communicate sustainability information to investors. Initially developed for SEC² filings and tailored to these specific requirements, the guidelines are also used globally.

Similarly, the Global Reporting Initiative (GRI) aids companies in communicating their impact on critical sustainability issues by developing global standards for sustainability reporting. The difference here is that the range of topics disclosed is more extensive, as these guidelines are not restricted to only financially material information but also encompass externalities and take a more general approach to the sustainability impact of companies. It is also designed with a globally applicable usage in mind. These efforts reflect a movement towards greater uniformity within industries, as they provide frameworks that companies can follow (Christensen et al., 2021).

Some countries and regions have decided to take CSR reporting practices and their importance even further. The EU also did so with its Non-Financial Reporting Directive (NFRD), which the European Parliament passed on 15 April 2014 (Directive 2014/95). This directive mandates publicly listed companies with over 500 employees and either assets exceeding 20 million EUR or sales over 40 million EUR per year to produce non-financial (CSR) reports, starting with the fiscal year 2017. Hence, the first reports under this directive were published in 2018 (Fiechter et al., 2022). The NFRD opts for a double materiality perspective, similar to the GRI, not only considering financially material information but also externalities and how their activities affect society and the environment (Christensen et al., 2021). This encompasses policies, main risks, and outcomes concerning environmental matters, social and employee aspects, human rights, anti-corruption, and diversity on the board of directors (Fiechter et al., 2022). The goal of the directive is to increase the transparency of the reported information across all member states and sectors (Directive 2014/95, recital 1), as well as setting an incentive for companies to engage in ESG-related activities (Directive 2014/95, recital 3) as neglecting the achievement of ESG related goals would be recognized and “punished” by investors and society. Additionally, the “disclosure of non-financial information helps the measuring, monitoring, and managing of undertakings’ performance and their impact on society” (Directive 2014/95, recital 3).

Despite these advancements, both guidelines and mandates leave considerable room for individual interpretation and framing. Voluntary guidelines allow companies to selec-

tively disclose certain aspects, tailoring their reports to reflect favorable characteristics while potentially omitting less favorable ones. Similarly, even mandatory directives can only partially standardize reporting due to the inherently qualitative nature of some data and the varied quantitative assessments that different companies might use. Also, it has been shown that mandates, in general, increase the use of boilerplate language and, hence, loss of report readability (Dyer et al., 2017). This might also be true for CSR reporting and perhaps be even more prevalent as the content is, by default, more qualitative, which could blur facts and talk even more. There is also a lack of detailed EU documentation or guidance on enforcement at the country level, which can lead to inconsistencies in how these regulations are applied (Fiechter et al., 2022).

This chapter highlighted why companies report on CSR-related topics, the complexity of CSR reporting, and approaches taken to harmonize reporting practices. While guidelines and mandates push companies towards greater transparency and accountability, they also require careful consideration to ensure that reports genuinely reflect the company’s impact rather than merely conforming to the letter of the law. In the following chapter, we will look at trends and determinants considering the extent and content of CSR reports and the possible effect of the NFRD before we take a closer look at the role of visuals, in particular pictures, in CSR reports. This will lay the foundation for my (picture) content analysis of ESG reports, focusing on the EU and Switzerland.

3. Trends in CSR Reporting

Building on the foundation set in the previous chapter, we will dive deeper into the practical implementation of CSR reporting and the factors that shape reporting outcomes. As we have seen, the central characteristic of these reports is the inherent heterogeneity in what companies report voluntarily. Reporting standards intend to reduce the problem, but as Christensen et al. (2021) highlight, companies’ adherence to these guidelines varies significantly. This is because, apart from standards, several factors influence reporting practices and outcomes. These include managerial incentives and other institutional arrangements. This implies that reporting also follows its own agenda and aligns with the previous theoretical discussions. Also, with the introduction of mandates aimed at enhancing uniformity, there are still challenges in aligning stakeholder expectations for transparency with actual business practices. These mandates often increase boilerplate language, as observed in various regulatory contexts like the K-10 disclosures in the US. This increase in generic content can reduce the specificity and readability of reports, potentially diluting the effectiveness of these disclosures in conveying clear and actionable information and obscuring specific data behind generic descriptions (Christensen et al., 2021; Dyer et al., 2017).

Furthermore, firms tend to focus predominantly on positive achievements in their reports while failing to critically engage with negative aspects, which can mislead the reader

² United States Securities and Exchange Commission

about the true state of CSR performance (Boiral & Henri, 2017). Other researchers made similar findings, pointing out that CSR communication can manipulate the reader, being decoupled from actions and attempting to benefit from a sustainable image without changing the company's actual practices (Siano et al., 2017). This aligns with my theoretical foundation that Legitimacy Theory and trying to create a positive picture of the company are key drivers in CSR reporting.

3.1. Determinants and Trends of CSR Reporting

When looking at the variability in CSR reporting and exploring its patterns, several determinants can be found that influence reporting behaviors. These include company size, industry characteristics, corporate performance, and litigation risk. These factors shape companies' strategic reporting behaviors and significantly influence how firms approach their CSR disclosures.

Company Performance: Companies showing a strong ESG performance are often more transparent and tend to disclose more numerical data and actual information that provides a clear view of their sustainability initiatives. Thereby providing a positive but accurate image of their performance. Similarly, a study of the disclosure of environmentally sensitive companies in China found that companies with better environmental performance report more on environmental matters and disclose more quantitative information (He & Loftus, 2014). In contrast, companies with relatively weaker performance may rely more on qualitative statements and use the reports to strategically project a positive image, thereby obscuring the reader (García-Sánchez & Araújo-Bernardo, 2020).

Influence of Company Size: The size of a company has a significant effect on reporting practices. It is, in fact, one of the determinants that have been consistently found to have a positive influence on sustainability reporting. Larger firms tend to have higher visibility and are assessed more thoroughly by stakeholders; thus, they are incentivized to extensively cover CSR topics. Christensen et al. (2021) highlight that the breadth of the issues addressed in CSR reports directly correlates with company size, with larger firms likely facing more media exposure and stakeholder pressure. Other papers support this finding by pointing out that the size of a company is a critical internal determinant that positively influences sustainability reporting, primarily because larger companies are more visible and, hence, more thoroughly investigated (Hahn & Kühnen, 2013).

Financial performance: Some researchers also found a positive correlation between a company's profitability and environmental disclosures, specifically suggesting that more profitable companies may have more resources to allocate toward disclosing environmental information. (Gamerschlag et al., 2011)

Industry Characteristics and Litigation Risk: The industry in which a company operates and the associated risks of that industry also influence reporting behaviors. High-polluting industries (environmentally sensitive industries)

report more extensively on CSR activities and information (Gamerschlag et al., 2011), possibly to mitigate potential backlash or regulatory controls. Similarly, Grougiou et al. (2016) found that companies in socially or environmentally sensitive industries – often referred to as “sin” industries – tend to report more on indicators that show socially responsible business practices. They argue that firms with significant litigation risks are more likely to use CSR reporting to build a responsible corporate image, potentially to mitigate backlash and minimize the direct or indirect costs of legal challenges. Another study found that companies in controversial sectors (here referring to socially sensitive industries) focus their reports more on community and people's achievements to enhance their legitimacy despite the negative impacts of their products or services (Byrd et al., 2017).

Selective reporting and Manipulation: As indicated at the beginning of this chapter, Guidelines, like the GRI, developed to increase transparency in CSR reporting, still must be critically assessed when it comes to their application. A Study focusing on companies that received the highest rating from the GRI initiative (A or A+ rating)³ discovered that there was still a significant underreporting of adverse events, violating GRI principles of balance, completeness, and transparency. This underreporting and the tendency to emphasize positive achievements over negative ones suggest that some firms attempt to maintain a positive corporate image rather than providing genuine transparency. Due to the inconsistencies in reported data and the qualitative nature of many GRI indicators, the challenges in the comparability of CSR reports have to be recognized even in the face of Guidelines (Boiral, 2013).

In this subchapter, we have examined the determinants influencing CSR reporting practices. Moving forward, we explore the specific impacts of reporting mandates using the NFRD as an example, exploring how they influence corporate behavior and the integrity of CSR disclosures.

3.2. Effects of a reporting mandate (NFRD)

The implementation of the NFRD has marked a significant shift in ESG reporting standards and practices in the EU, with the aim to enhance transparency and comparability across firms. There are several findings that indicate its (partly) success. A study among Spanish firms from 2013-2018, for example, found that there has been a significant increase in companies adhering to the GRI standards, indicating an improvement in transparency directly influenced by the directive. (García-Sánchez & Araújo-Bernardo, 2020)

A more detailed analysis across a broader spectrum of EU companies by Fiechter et al. (2022), studying the effects of the NFRD, noted an increase in reporting transparency beginning in 2014, the year the resolution was passed. With their

³ Levels A+ and A are the most demanding and complete levels of use of this reporting system. For instance, Level A assumes that the sustainable development report considers each core and sector supplement indicator. Level A+ assumes, in addition, that the report is audited by a third party (GRI, 2006; Boiral and Henri, 2017, p. 5)

research focusing on the years 2011 to 2018, they highlighted that no significant increase in transparency was observable before 2013, suggesting that this effect is also attributable to the NFRD catalyzing these improvements.

The distribution of the mandate's impact was not consistent across all dimensions of CSR. Although there was a notable rise in CSR activity, as measured by the combined social and environmental scores, the effect was only statistically significant for the social score when analyzed independently (Fiechter et al., 2022). The limited impact on the environmental score could be attributed to longer implementation times and the long-range nature of environmental initiatives. Additionally, the most substantial enhancements were observed in companies categorized as "high exposure" – those with lower levels of transparency and activity prior to the mandate (lower than the median of observed countries before 2014). These firms also showed improvements in their environmental activities by 2018, indicating that the NFRD has been particularly effective in enhancing reporting and CSR practices among firms that were initially less compliant (Fiechter et al., 2022).

Additionally, the directive seems to positively affect investments into CSR infrastructure. Companies with a higher exposure prior to the NFRD engaged more in establishing an internal CSR Committee and launching social initiatives, often followed by an increase in the ESG Score (Fiechter et al., 2022).

Despite these positive changes, they acknowledge that there is no detailed EU documentation or guidance on the enforcement of these standards at the country level, which could impact the consistency and effectiveness of the NFRD's implementation across member states (Fiechter et al., 2022).

3.3. Preliminary Reflections on the Role of the Visuals in ESG Reports

While the NFRD has improved transparency in CSR reporting, the usage of pictures was not included in the analysis. However, the use of images is interesting. Studies highlighted that pictures in the context of CSR reports could be used for greenwashing (Chong et al., 2019) by overshadowing less favorable environmental or social impacts, potentially leading stakeholders to misinterpret a company's actual practices and performances.

Moreover, other studies discuss how boilerplate language, which increases in the face of mandates, can be used for greenwashing in CSR reports (Christensen et al., 2021). While they do not specifically mention visuals, images can amplify this effect, and if noninformative, could be seen as a type of boilerplate itself. Visuals can emphasize positive aspects or downplay negative ones, subtly shaping stakeholder perceptions and potentially enhancing the impact of boilerplate content.

In addition, when it comes to greenwashing, it has been shown that passive greenwashing, in contrast to active greenwashing⁴, is perceived as less severe and induces smaller neg-

ative effects. In addition, communications compared to action greenwashing has been identified to have a smaller negative effect in the scope of passive greenwashing. Thus, passive greenwashing through communications has been found to be the least severe form of greenwashing (Gatti et al., 2021). The usage of pictures is most likely to be attributed to that category and is perhaps even more subtle, making it important to understand the effects and intentions with which pictures are used.

We preliminarily highlighted the role visuals might play in CSR reporting, indicating that while they may enhance the clarity of reports, they could also be used to create a favorable (false) image. The next chapter will discuss the psychological and communicative effects of visuals, examining how they influence perception. Following that, we will review existing research on the uses of visuals in CSR reports to establish a foundation for my analysis.

4. The Significance of Visuals in Perception

4.1. Impact of Visuals on Perception and Cognition

The human brain processes visuals in evolutionarily older areas of the brain (Harper, 2002), indicating their importance in shaping our perception and cognition. We also pay more attention to visuals than we do to text (Tversky, 1974) - often at a subconscious level where their subtle influences bias our perception without explicit awareness (Posner et al., 1976). In general, visuals help in memorizing and have a more powerful place in memory (Vasue & Howe, 1989), but when we process and focus on too many visuals, it can potentially lead to cognitive "overloading" and, thus, distract from key information (Tversky, 1974).

Visuals can play a crucial role in forming aesthetic appeal, which in turn influences how individuals perceive information. Research has demonstrated that aesthetics have the power to generate a positive impression, making them highly valuable in the realm of corporate communication (Legendre et al., 2020). A study conducted by Townsend and Shu (2010), provides evidence of this phenomenon. In the study, investors were tasked with evaluating the value of a fictitious company based on its annual report, with the level of aesthetic elements (such as the quantity, size, and color of pictures/images) varying across different groups. They observed that aesthetics impacted the judgment of investors about the company to the same extent as financial metrics, revealing that even experienced stakeholders can be influenced by the design choices of annual reports. Another interesting observation is that when investors have been made aware that aesthetics could influence their valuation⁵, the effect diminishes (Townsend & Shu, 2010), strengthening the argument that the influence of visuals is often subconscious.

⁴ formation, while passive greenwashing refers to covering up information (Gatti et al., 2021)

⁵ By asking the investors beforehand how much they think that aesthetics influence their decision together with other factors like profitability etc.

⁴ Active greenwashing refers to companies intentionally creating false in-

These findings are crucial for understanding how companies might leverage photographs to enhance their corporate image.

4.2. Strategic Role of Visuals in Corporate Narrative Construction

There are two primary purposes that visuals can serve in reporting. On the one hand, they can enhance written information by visually representing “reality” (describe). On the other hand, they have the power to create and convey their own message (construct) (Davison, 2015). Visuals can construct complex narratives, which are vital in shaping an organization’s identity and influencing how stakeholders perceive it (Preston et al., 1996). However, because visuals have the capacity to create and manipulate reality, it is important to approach them critically.

Davison (2010) discusses an example of the complex messages transported through images in the context of Business portraits. They serve to provide a human face to the organization, thereby establishing a connection with the company and increasing investor confidence and customer loyalty. They are also used to “brand” the top executives and, therefore, the company as reliable, competent, and ethical (Davison, 2010). Repeating a theme throughout a document enhances the ability of visuals to construct. As they become more familiar, recognition and retention are increased, favoring the perception of the company (and their practices, depending on the theme) (Davison, 2008).

Due to their potent ability, the use of visuals is not without concerns. Cho et al. (2009) express that visuals could potentially mislead stakeholders by overshadowing or distracting from the actual content. Their study showed that adding pictures to a text-only representation on corporate websites leads to a significant increase in perceived social responsibility without altering the actual information being represented. This change in perception, achieved solely by adding pictures, could give rise to potential greenwashing. This effect of pictures can likely also be extended to communication via CSR reports. With this, we now continue by looking at what has already been found out about the usage of pictures in ESG reports.

5. Insights into the Use of Visuals in CSR Reports

The previous chapters give reasons to believe that visuals are not only integrated into CSR reports for aesthetic or explanatory reasons but also serve a crucial role in the preparation of reports (Invernizzi et al., 2022). Over the last few years, there has been an increase in the number of pictures used in CSR reports (Chong et al., 2019). Despite these developments, this area has only been touched upon by a few studies (Davison, 2015).

5.1. Strategic Use of Visual Content in CSR Reports

From the viewpoint of legitimacy theory, organizations use CSR reports to strengthen their public image and mar-

ket position. Images seem to be a central part of this strategy. Rämö (2011) highlights that photographs in CSR reports serve to support the narrative of responsibility and ethical behavior, showcasing the commitment to societal values and integrity. Similarly, Invernizzi et al. (2022) conclude that these images paint a picture of competence and achievement in the economic and CSR realm.

This attempt to portray competence and achievement leads to a significant bias towards predominantly using positive images and creating an idealized view of the company, downplaying negative effects (Boiral, 2013) and, as indicated in the previous chapters, could mislead the reader about the actual company’s performance. Particularly striking is that the images almost occupy the same amount of space as the textual content in CSR disclosures (Boiral, 2013). This bias of visuals and the large space they occupy can distance the report from actual sustainability performance, potentially enhancing the company’s image at the expense of transparency. Furthermore, it has been observed that less sustainability-driven companies additionally use generic, non-specific images, symbolically aligning themselves with sustainability themes and enhancing their image. This approach can be seen as a form of greenwashing, as the portrayed and actual sustainability performance and attitude don’t align (Hrasky, 2012).

5.2. Analysis of Visuals in CSR Reports

Regarding the content of the pictures being used, the social dimension is the most frequently depicted, followed by the environmental. Notably, Invernizzi et al. (2022) found that on average, 2.38 photos per report depict an environmental theme and 7.47 a social one. Another finding from their study is that regarding readability and perceived legitimacy, the optimal number of pictures per page is about one per page.

Another study focuses on the size and color of the pictures being used while differentiating between standardized (report on all GRI indicators) and non-standardized (do not report on all GRI indicators) reports. Less standardized reports tend to use larger, more colorful, and overall, more images, potentially compensating for the lack of substantive information (García-Sánchez & Araújo-Bernardo, 2020), further complementing the viewpoint that companies use images strategically to form impressions.

Following an industry-specific focus, Chong et al. (2019) find that companies from “sensitive” industries in New Zealand, such as energy, mining and construction, property, agriculture, fishing, and transport, use more pictures depicting an environmental theme, again possibly to enhance the perception of their environmental performance and to address stakeholder expectations.

Together with the preceding chapters, these insights highlight the importance of understanding what factors determine companies’ use of images and the rationale behind this behavior.

In the following, I will concretize the research question and look at the data used for the analysis, the methodology

applied, and the research design before evaluating the results.

6. Methodology and Data

6.1. Data Collection

For the analysis, I gathered two types of data. One is the information about the pictures in CSR reports, which I will call “primary data” in the following, and the other is supplementary data encompassing company-specific metrics on a yearly basis (e.g., about the industry, financial metrics, ESG-related information, and company characteristics), this will be called “secondary data”.

The companies I have initially considered include all companies from the STOXX Europe 600 (EU companies) that were part of the index in all years between 2011 and 2020, and all companies from the index SMI Expanded (Swiss companies)

To gather the primary data in a format suited for analysis, I used the Google Cloud Vision AI (Google, 2024) to first detect each picture from the reports and then automatically create a list of labels (in most cases consisting of 10 unique labels) that describe its contents. The information gathered was then stored on a “company-year-picture level”. What is to be noted is that I only considered sustainability reports published separately from the annual financial report to be able to distinguish between financial and ESG information clearly and automatically.

The secondary data was retrieved from the LSEG Workspace (former Refinitiv), comprising a broad range of metrics for my desired financial and ESG dimensions. I further enriched this (secondary) data by adding dummy variables indicating whether a company operates in an environmentally or socially sensitive industry and is operating close to its customers. This classification was done at the industry level (ICB Codes). Industries considered environmentally sensitive were chosen based on the classification of Branco and Rodrigues (2008) but adjusted to the ICB Codes and adjusted to include airlines. They encompass the following: Airlines, Construction and Material, Oil, Gas and Coal, Electricity, Gas, Water, and Multi utilities, and Basic Materials. Similar socially sensitive industries are based on the classification from Brammer and Millington (2005), again adjusted to the ICB Codes, and encompass Pharmaceuticals, Tobacco, Alcohol and Soft drinks, Defense, and Basic Materials. Industries operating close to their customers are again based on the classification used by Branco (Branco & Rodrigues, 2008), adjusted to the ICB Code classification, and additionally considering media, they encompass Telecommunications, Consumer Products and services, Media, Retailers, Consumer Staples, Electricity, Gas, Water, and Multi utilities.

A complete overview of the considered metrics, their explanation, and their availability can be found in Appendix A.

6.2. Picture Classification

I propose a unique approach that is different from manual content analysis, as I first leverage the Cloud Vision AI

(Google, 2024) to automatically detect and describe the images and then use a Natural Language Processing (NLP) model to embed these descriptions in a vector space suitable for clustering methods, which are then applied. This approach enabled me to analyze a total of 11.455 pictures.

The list of labels describing each picture for every company every year, created by the Cloud Vision AI, constitutes the basis for my picture classification approach. I treated this list as a sentence for the NLP to be able to capture the semantic meaning of each label in the context in which it is being used. The model that was opted for is MPnet, which is built upon the architecture of BERT and published in cooperation with Microsoft Research. I chose this model as it effectively addresses BERT’s limitations in handling token dependencies and position information, leading to significantly improved performance on a variety of language understanding tasks (Song et al., 2020). Using MPnet, I created word embeddings for each Label List; after some tryouts, I identified that a reduction in the dimensionality was necessary to better capture the semantic similarities between the embedded Lists. Here, I identified that a Principal Component Analysis (PCA) reduction (Wang, 2019) of 0,7 delivered the best results in reducing the high dimensionality while keeping enough information to properly distinguish their semantic meanings.

To create clusters based on these embeddings, I used k-means clustering. To find the right number of clusters, I followed two steps. In the first step, I used the Elbow Method (Humaira & Rasyidah, 2018) to define the range where the optimal number of clusters lies. The analysis showed that the inflection point is between 5 and 30. In the next step, I only considered this range, so I looked at the Silhouette Score and Davies Bouldin Score (Yilmaz et al., 2007) for each k in that specified range as a measurement for cluster quality. I identified the three clusters with the best results based on these measurement scores and further assessed them through sample testing. After this assessment, I opted for 15 initial clusters⁶. The sample testing gave insights into further manual adjustments that could be made via the consolidation of clusters and reallocation through keywords. The consolidation took place after the reallocation and can be seen in Figure 1, together with the initial and final clusters and their respective names. The reallocation was done in three steps.

First, I created a new Cluster containing images showing renewable energy generation. I identified that if a label in the label list contains the words “wind” or “solar” while not being in the defined exclusion list⁷, it can be allocated to this new category.

In the next step, I identified 4 clusters (2, 8, 9, 13) containing images depicting modes of transportation that belong to Cluster 9. Here, I also identified a list of keywords⁸ and an exclusion list⁹ where if a label contains a word in the keyword

⁶ Silhouette Score: 0,162; Davis Bouldin Score: 1,865

⁷ Exclusion List for Renewable Energies: [‘window’, ‘windshield’, ‘windbreaker’, ‘windscreen’, ‘wind instrument’, ‘wind wave’, ‘wind’]

⁸ Keywords Transportation: [‘boat’, ‘transport’, ‘air’, ‘ship’, ‘rail’, ‘rolling stock’, ‘naval’]

⁹ Exclusion List Transportation: [‘boats and boating-equipment and sup-

list and is not in the exclusion list, the label gets reallocated. The rest of Clusters 8 and 9 were kept, and Clusters 2 and 13 were further refined.

Cluster 2 now only had a few pictures left depicting nature, so I allocated the rest of the Cluster to the Cluster for nature images (Cluster 4). Similar to the approach applied to Cluster 2, I also used a keyword-based filtering process for Cluster 13. Specifically, if the label list contained one of the keywords ‘urban design’, ‘city’, or ‘skyscraper’, the respective images were allocated to the cluster depicting urban design. The remaining images in Cluster 13 were then reallocated to Cluster 4, which contains images of nature.

Following this and the consolidation, I further summarized the clusters into three categories: Environmental, Social, and Corporate Excellence. This final consolidation can be seen in Figure 1.

6.3. Company Matching

I opted for a direct mapping approach to create comparable groups between the EU and Switzerland. In a first step, I filtered out companies that reported (published a separate sustainability report) less than four times between the years 2014 and 2020. In the next step, I performed a full join on all companies from the EU and Switzerland that fulfilled the criteria based on their two-digit ICB (Industry Classification Benchmark) Codes. This allows companies to be matched to more than one counterpart from another region.

Following this, I filtered these matches based on their percentage difference in revenue. Only matches with a difference smaller than 20 percent were kept, and the companies that resulted from these matches were defined as my sample. Table 1 shows the number of companies reporting in each year between 2013 and 2020, differentiated by region and in total. Appendix B shows the distribution of the companies between the two-digit ICB Code names, also differentiated by region and total.

7. Analysis and Results

7.1. Research Question and Design of the Study

In my analysis, I want to explore the amount and type of pictures companies use in their CSR Reports, whether there are determinants that influence this decision, and test whether I can see a difference between the usage in the EU and Switzerland, potentially due to the introduction of the NFRD.

I follow an exploratory approach in my analysis, but I expect to see several things based on the literature. Companies that are either in a sensitive industry or have a comparably worse ESG performance might increase the number of pictures being used to create a positive image of the company by directing attention more to the pictures and not the actual

information. Similarly, I expect companies to use a higher ratio of picture themes in areas where they might face the most attention, either through their industry sensitivity or poor sustainability performance. Lastly, as the introduction of the NFRD led to increased reporting transparency (Fiechter et al., 2022), I expect to see the same regarding picture usage. I define Transparency in this setting as a decrease in the number of images being used.

I will structure my analysis in two separate parts. In the first part, I don’t consider the region (EU and Swiss) where a company is listed. I utilize the whole dataset to analyze the determinants of picture usage in ESG reporting via descriptive statistics and regression models. In the second part, I examine the differences between the two regions to determine the treatment effects of introducing a reporting mandate (NFRD). Here, I will start with an overview using descriptive statistics and graphs. Then, I will test whether these findings are significant using statistical tests and a difference-in-difference regression approach, outlined in detail in the analysis. When a variable I use is inside the expression “log(...)”, I use the natural logarithm of this variable for the analysis.

7.2. Determinants of Picture Usage in CSR Reports

7.2.1. Analysis of Photographs by Page Count

I start my analysis by looking at the average number of pictures being used. In Table 2 we see that companies on average use 0.506 pictures per report, while companies in the EU (0.576) tend to use more pictures than companies in Switzerland (0.400), thus giving me reason to believe that region is a first determinant of picture usage. Differentiating the usage by year, shown in Figure 2, shows that the difference between the regions remained roughly constant until 2017-2018. From this point onwards, Swiss companies seem to have closed this gap, which will later be explored in section 7.3.

To test whether the region is a significant determinant and find further ones, I used a backward elimination (linear) regression approach. I started with a linear regression including all variables¹⁰ that might influence “Photograph by Page Count”. I stepwise eliminated variables showing the highest P-Values, which increased the adjusted R-squared when left out until no further increase in the adjusted R-squared was possible without eliminating potentially significant variables. Ultimately, I ensured that all sensitivity indicators (ENV_Sensitive, S_Sensitive, E_X_S) were in the model when at least one was considered after the final elimination to ensure I captured their dependencies. The result of the final regression can be seen in Table 3. Here, we find three significant variables. As indicated in Figure 2, we see that the

¹⁰ These include: Log(Revenue_MEUR), Log(Employees), Log(ESG_Score), Log(GRI_Reporting_Score), Log(CO2_Emissions_Rev), ROA, Avg_Board_Ten, Log(Total Assets_MEUR), CSR Audit, ENV_Sensitive, S_Sensitive, E_X_S, ENV_Controversies, ENV_Initiative, EU, Customer_Proximity, CSR_Infrastructure, ESG_Controversies_Score

plies’, ‘mode of transport’, ‘transport hub’, ‘pipeline transport’, ‘long hair’, ‘public transport’, ‘chair’, ‘hair’, ‘window’, ‘stair’, ‘stairs’, ‘dairy’]

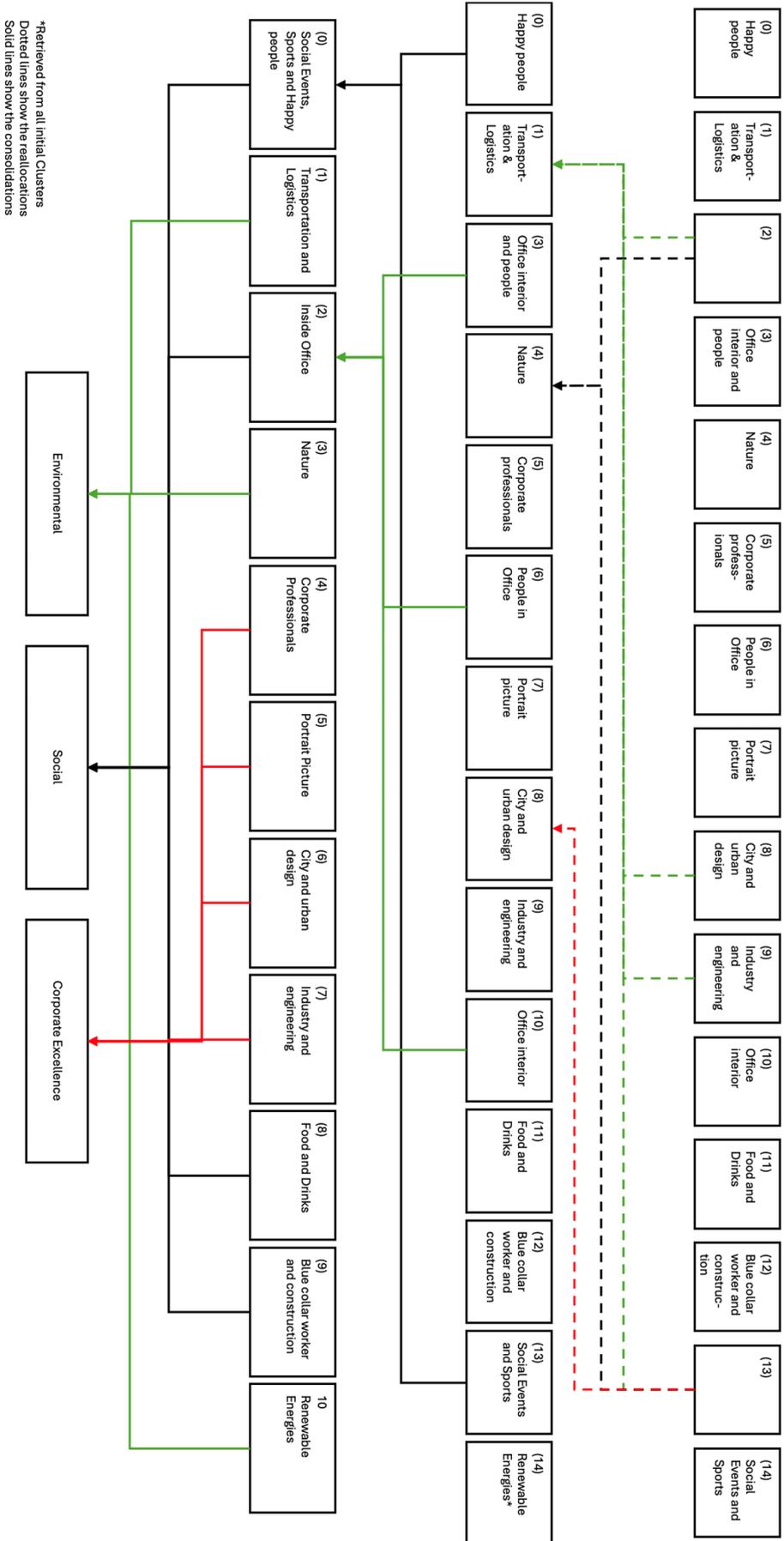
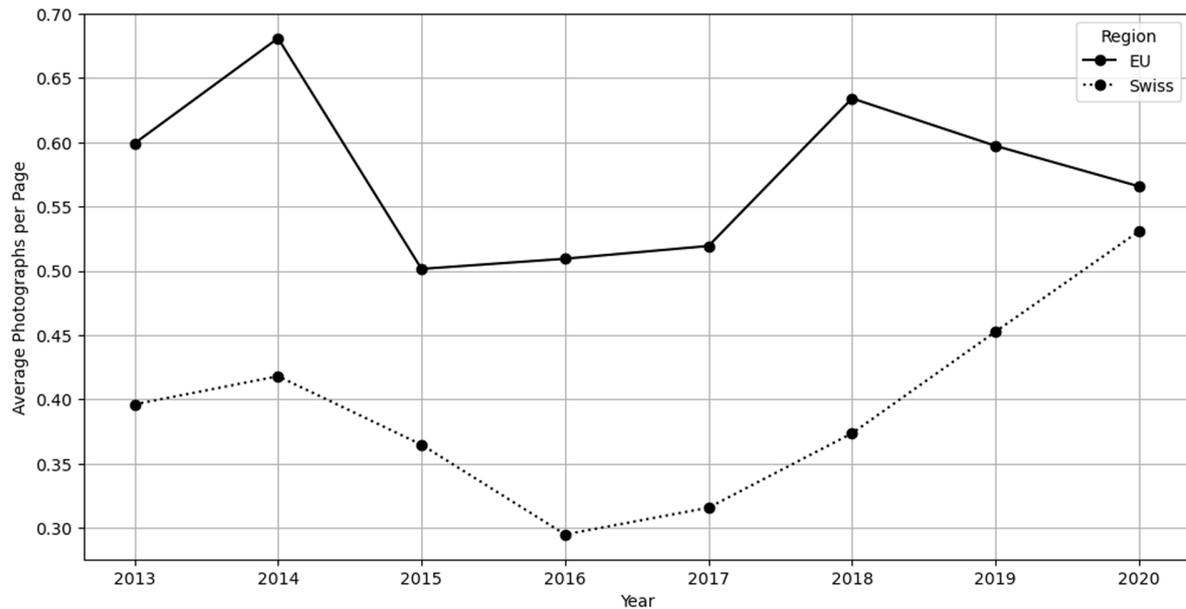


Figure 1: Overview of Reallocation and Consolidation of Picture Categories (Source: Own Illustration)

Table 1: Number of Companies considered in the sample (Each year and in Total)

Region	2013	2014	2015	2016	2017	2018	2019	2020
EU	31	33	31	35	36	35	35	32
Swiss	10	15	19	20	25	29	32	29
Total	41	48	50	55	61	64	67	61

**Figure 2:** Pictures per Report (Differentiated by Region and Year) (Source: Own Illustration)**Table 2:** Photograph by Page Count (Differentiated by Region)

EU	Swiss	Total
0.576	0.400	0.506

region influences the number of pictures, namely that companies in the EU use more images, shown by the variable EU (beta: 0.105; P-value: 0.022). In addition, we find two further significant variables. The extent to which a company discloses GRI indicators, meaning a broader and more extensive coverage of CSR topics, has a negative effect on the number of pictures being used (beta: -0.14; P-value: 0.000), a one percent change in the “GRI_Reporting_Score” decreases the number of pictures being used by 0.14. The third significant variable is the CO2 intensity (“Log(CO2_Emissions_Rev)”), measuring the amount of CO2 (in Tn’s) emitted for every million in revenue. Here, we see a positive influence (beta: 0.09; P-Value: 0.000), meaning that a one percent increase in the CO2 emitted for the same revenue increases the number of pictures used by 0.09.

7.2.2. Analysis of Picture Contents

Switching the focus to the content of the pictures, I first investigated the distribution of the three picture categories derived in 6.3 (Environmental, Social, and Corporate Excellence), by region and in Total, which can be seen in Table 4. I calculated the final averages by averaging the individual averages of each category for each report. Here, we see no apparent difference between the regions, and the dominant theme is the social dimension (with almost half of the pictures), followed by Corporate Excellence and, lastly, Environmental themes. In line with the analysis in 7.2.1, I followed a backward elimination approach to find variables that influence the distribution of picture categories within a report. I conducted three separate (linear) regressions with the percentage of each picture category as the dependent variable. For each regression, I again started with all variables¹¹, which I assumed to potentially influence the distribution of picture categories.

¹¹ These include: Log(Revenue_MEUR), Log(Employees), Log(ESG_Score), Log(GRI_Reporting_Score), Log(CO2_Emissions_Rev), ROA, Avg_Board_Ten, Log(Total_Assets_MEUR), CSR_Audit, ENV_Sensitive, S_Sensitive, E_X_S, ENV_Controversies, ENV_Initiative, EU, Customer_Proximity, CSR_Infrastructure, ESG_Controversies_Score, E_Score, S_Score

Table 3: Regression on Photograph by Page Count

Variable	Photograph by Page Count		
	Coefficient	Std.	VIF [†]
Intercept	0.4514	(0.061)	427.65
CSR Audit	-0.0398	(0.074)	1.58
EU	0.1050*	(0.045)	3.10
Customer_Proximity	0.0686	(0.053)	1.18
Env_Sensitive	0.0033	(0.098)	1.27
S_Sensitive	0.1601	(0.139)	1.90
E_X_S ¹	-0.1200	(0.171)	2.57
Log(ESG_Score)	0.0375	(0.022)	1.35
Log(GRI_Reporting_Score)	-0.1412***	(0.029)	1.14
Log(CO2_Emissions_Rev)	0.0879***	(0.021)	1.55
ROA	0.0129	(0.020)	1.06
Adjusted R ²		0.186	

*** p < .001, ** p < .01, * p < .05

[†]Variance Inflation Factor (VIF): I included the VIF to test for Correlation of independent variables

Notes: N = 317 company-year observations, F-statistic 11.02, Std = Standard deviation of Intercepts, Robust standard error were used, all continuous variables have been standardized before the regression

Table 4: Picture Categories by Region and in Total

Region	Environmental	Social	Corporate Excellence
EU	21,89%	44,51%	33,60%
Swiss	21,64%	43,59%	34,77%
Total	21,80%	44,17%	34,03%

I eliminated all highly insignificant variables stepwise when increasing the adjusted R-squared by being left out. Apart from ensuring that all sensitivity indicators are considered when one of them is in the final regression, I also adjusted the final regressions to include the same variables for each regression for better comparability by adding them back into the regression when significant for another regression. Again, I tested for multicollinearity by calculating the VIF scores, which can be seen in Appendix C. The results of these regressions can be seen in Table 5. Here, we see several significant variables and interplays between the categories. Starting with a simple effect, we see that companies taking environmental initiatives (“ENV_Initiative”) tend to use more environmental-themed pictures (beta: 0.05; P-value: 0.032). Next, we see that Revenue (beta: 0.07; P-value: 0.006) has a positive effect, and the number of Employees (beta: -0.060; P-value: 0.016) has an adverse impact on the number of environmental pictures being used. Another variable significant for only one category is the ESG Score in the Social category. Here, we see that a higher ESG Score is indicative of higher usage of pictures with a social theme (beta: 0.06; P-value: 0.019). We also see that socially sensitive in-

dustries (“S_Sensitive”) tend to use fewer pictures depicting environmental themes (beta: -0.11; P-value: 0.018). While the variable is insignificant for the other picture categories, it could be reasoned that these companies use a higher percentage in another category. Looking at the coefficients, these presumably use more socially related pictures (0.001 vs. 0.096), but this effect cannot be statistically verified. Now, the focus is switched to the variables significant for both Environmental and Corporate excellence. We see that the CO2 intensity of revenue generation (“Log(CO2_Emissions_Rev)”) is significant and increases the percentage of Environmental pictures (beta: 0.05; P-value: 0.000) while decreasing the percentage of pictures used from the Corporate Excellence category (beta: -0.06; P-value: 0.000). A similar effect can be seen for companies being both environmentally and socially sensitive (“E_X_S”). For these companies, we also see an increase in the percentage of environmental pictures (beta: 0.276; P-value: 0.001) and a decrease in the percentage of pictures depicting Corporate Excellence (beta: -0.236; P-value: 0.032). Interestingly, the single variable of being environmentally sensitive (“ENV_Sensitive”) has the contrary effect, decreasing the Environmental pictures (beta:

Table 5: Regression on Picture Contents

Variable	Environmental		Corporate Excellence		Social	
	Coefficient	Std.	Coefficient	Std.	Coefficient	Std.
Intercept	0.246	(0.030)	0.354	(0.042)	0.410	(0.044)
CSR Audit	0.023	(0.036)	0.014	(0.043)	0.009	(0.045)
ENV_Sensitive	-0.118***	(0.027)	0.185***	(0.051)	-0.067	(0.048)
S_Sensitive	-0.106*	(0.045)	0.001	(0.072)	0.096	(0.075)
Customer_Proximity	-0.021	(0.033)	-0.018**	(0.032)	0.129***	(0.035)
E_X_S	0.276**	(0.082)	-0.236*	(0.110)	-0.040	(0.108)
ENV_Initiative	0.053*	(0.025)	-0.024	(0.024)	-0.029	(0.027)
Log(Revenue_MEUR)	0.075**	(0.027)	-0.036	(0.026)	-0.038	(0.029)
Log(Employees)	-0.060*	(0.025)	0.022	(0.026)	0.038	(0.027)
Log(ESG_Score)	-0.030	(0.025)	-0.031	(0.026)	0.061*	(0.026)
E_Score	0.016	(0.020)	0.031	(0.024)	-0.047	(0.024)
S_Score	-0.017	(0.020)	0.036	(0.022)	-0.019	(0.022)
ROA	-0.008	(0.010)	-0.007	(0.012)	0.015	(0.012)
Avg_Board_Ten	-0.014	(0.010)	0.018	(0.013)	-0.004	(0.013)
Log(CO2_Emissions_Rev)	0.054***	(0.014)	-0.058***	(0.014)	0.003	(0.014)
Adjusted R ²	0.105		0.108		0.096	

*** p < .001, ** p < .01, * p < .05

Notes: N = 342 company-year observations, Std = Standard deviation of Intercepts, Robust standard error were used, all continuous variables have been standardized before the regression

-0.118; P-value: 0.000) while increasing the Corporate Excellence pictures (beta: 0.184; P-value: 0.000). Lastly, we see a significant effect for the closeness of a company to its customers (“Customer_Proximity”). For this variable, we see a comparably strong and positive effect on the percentage of Social pictures being used (beta: 0.129; P-value: 0.000) and a smaller negative effect on Corporate Excellence pictures (beta: -0.018; P-value: 0.001).

7.3. Comparison of Picture Usage between EU and Switzerland

To analyze the difference between the EU and Switzerland, we first look at the number of pictures used per report page (“Photograph by Page Count”). In Figure 2, we saw that EU companies generally use more pictures per report page over the whole period. Still, we see that during the last periods (2017-2020), this difference becomes smaller, giving me reason to believe that around the time of the introduction of the NFRD, a change is happening in this regard. To better visualize this potential effect, I index the lines. Initially assuming that the effect starts either right after it was passed (2015 – indexed in 2014; chart can be seen in Appendix D) or in the first year the NFRD applies (2017 - indexed in 2016; chart can be seen in Appendix E), I didn’t find unusual changes after the indexed year. However, indexing the 2018 values shows us two interesting things, which can

be seen in Figure 3. First, after the year 2018, the change in the number of images used per report page was positive for Switzerland while negative for the EU. Second, we see that before 2018, the trend of the number of pictures used per report page was almost parallel for both regions. This initial (almost) parallel trend with a clear change happening in 2018 indicates that there might be a treatment effect, induced by the introduction of the NFRD, on the number of pictures being used. Figure 4 further differentiates into sensitive and non-sensitive companies (either environmentally or socially sensitive). This graph indicates that this effect (both the positive trend for Swiss companies and the negative trend for EU companies) might be stronger for socially or environmentally sensitive companies. After verifying the parallel trend assumption, I used a two-stage difference in difference regression approach to test for these effects. The validation was conducted using the approach outlined by Riveros-Gavilanes (2023). These regression results can be found in Appendix F. In the first stage (Model 1), I created three binary variables: Treatment, indicating whether the company at some time received the Treatment (EU Companies), Post indicating whether the observation was before (<2019 – “0”) or after (>=2019 – “1”) the supposed Treatment, and their interaction “Treat_X_Post” indicating whether the company is treated (EU) and observed after the supposed treatment year, my main variable of interest. In the second stage (Model

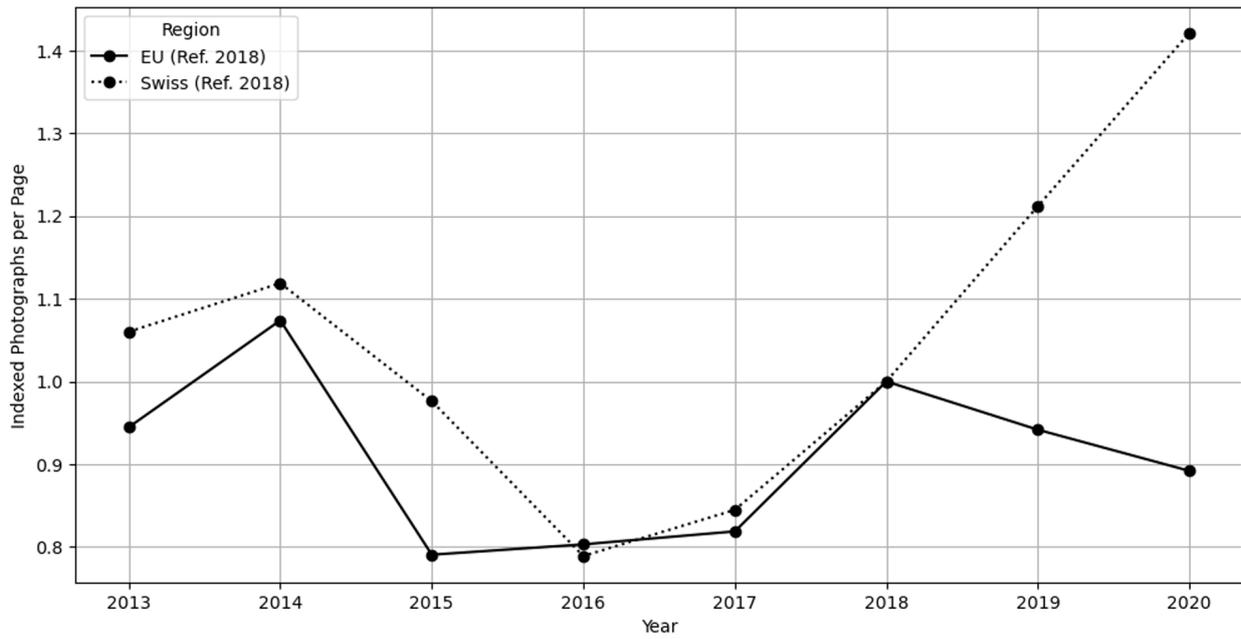


Figure 3: Pictures per Report Page in the EU and Switzerland (Indexed in 2018) (Source: Own Illustration)

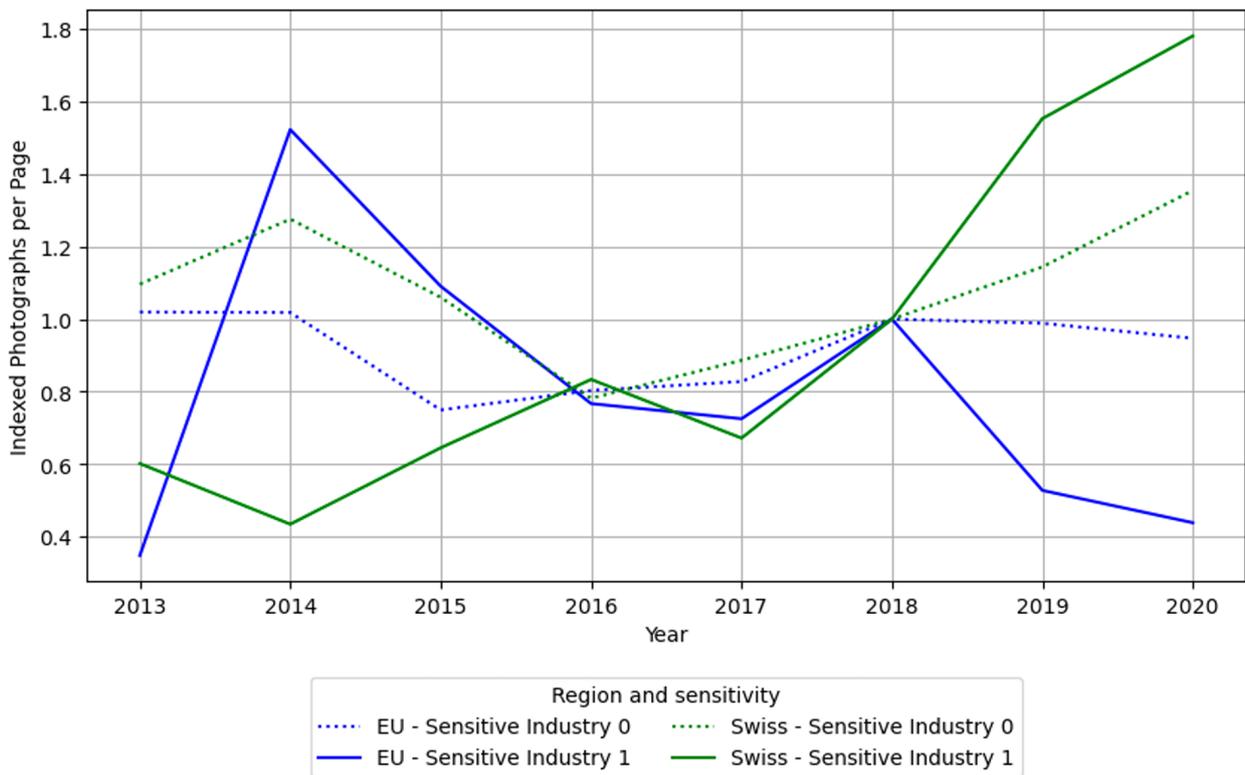


Figure 4: Pictures per Report Page separated by Industry sensitivity (Indexed in 2018) (Source: Own Illustration)

2), I further added the variable “Any_Sensitivity,” indicating whether the company is either environmentally or socially sensitive and its interaction with all the previous variables. In this setup, my main variables of interest are “Treat_X_Post” and the triple interaction term “Treat_X_Post_X_Sensitivity”.

Beginning with Model 1, my primary variable of interest, “Treat_X_Post,” shows a negative effect (beta: -0.185) on “Photograph by Page Count”. This suggests that post-2018, assuming a parallel trend in the absence of the treatment, EU companies used fewer images than anticipated. While

Table 6: Difference in Difference Regression on Photograph by Page Count

Variable	Model 1			Model 2		
	Coefficient	Std.	P-Value [†]	Coefficient	Std.	P-Value [†]
Intercept	0.294	(0.049)	[0.000]	0.328	(0.053)	[0.000]
Treatment	0.272***	(0.054)	[0.000]	0.226***	(0.060)	[0.000]
Post	0.192**	(0.068)	[0.011]	0.151*	(0.076)	[0.062]
Treat_X_Post	-0.185*	(0.090)	[0.056]	-0.106	(0.098)	[0.299]
Any_Sensitive				-0.193	(0.105)	[0.002]
Treat_X_Sensitive				0.222	(0.139)	[0.116]
Post_X_Sensitive				0.188	(0.168)	[0.426]
Treat_X_Post_X_Sensitive				-0.563*	(0.260)	[0.064]
Adjusted R ²		0.070			0.077	

*** p < .001, ** p < .01, * p < .05 (when using non robust standard errors)

[†]P-values when using robust standard errors

Notes: N = 407 company-year observations, Std = Standard deviation of Intercepts, all continuous variables have been standardized before the regression

this effect is statistically significant at conventional P-value thresholds (0.039), it becomes marginally non-significant (P-value: 0.056) when adjusted for robust standard errors. I additionally used robust standard errors as testing the model indicated potential heteroscedasticity. Consequently, I cannot definitively dismiss the null hypothesis of no treatment effect, though the evidence strongly hints at a potential effect. In Model 2, “Treat_X_Post” represents the impact on non-sensitive EU companies and shows no significant effect under any P-value conditions. However, the triple interaction term, “Treat_X_Post_Sensitive,” which measures the impact on sensitive industries within the EU compared to Switzerland, initially appears significant at standard P-value thresholds (0.031) but, again, is marginally non-significant (P-value: 0.064) with robust standard errors. Despite being unable to reject the null hypotheses in both models, I observe strong evidence of a treatment effect, particularly concentrated among sensitive companies.

Table 7: Difference of the number of pictures per report page between 2018 and 2020

	Sensitive Companies	Non-Sensitive Companies	Total
EU	- 0.424	- 0.033	- 0.069
Swiss	+ 0.265	+ 0.135	+ 0.157

To quantify this effect in descriptive terms, Table 7 shows the average “Photograph by Page Count” difference between 2018 and 2020 in the EU and Switzerland, further differentiated by sensitivity. Here, we see the largest difference between the EU and Switzerland in Sensitive companies. While companies in the EU, on average, reduced the number of

pictures per report by 0.42, the companies in Switzerland increased the number of pictures used by 0.27, resulting in a change in their difference of 0.69. Comparing this to the averages over the whole period for each region depicted in Table 2 (0.576 for the EU and 0.400 for Switzerland), this descriptive approach indicates that the magnitude of the potential Treatment effect is considerable on an absolute level.

8. Discussion and Limitations

8.1. Discussion

In this section, I relate my findings to the theoretical framework discussed earlier and offer potential explanations for results that are not directly supported or contradicted by existing theory. To recap, I expected to see companies in sensitive industries or with comparably lower ESG performance using pictures more frequently, possibly creating a more favorable image by redirecting attention away from less positive information. Further, I anticipated that companies would use a higher ratio of pictures depicting a theme from the area where they face greater scrutiny or have poor sustainability performance. Finally, knowing that the introduction of the NFRD increased reporting transparency, I expected to see a corresponding decrease in the use of pictures, interpreting increased transparency as reduced reliance on images.

8.1.1. Frequency of Picture Usage in CSR Reports

My analysis indicates three variables that influence the frequency of image usage. While broader characteristics such as ESG scores or industry sensitivity (like initially expected) did not directly correlate with the frequency of images being used, significant effects were found for CO2 Emissions per revenue. This might suggest that concrete and measurable performance metrics influence image use rather than general sensitivity or ESG performance, possibly due to their direct

disclosure and visibility in CSR reports. CO₂ emissions are central in CSR reporting and are easily understandable even by less experienced readers. Firms with higher CO₂ intensity than their peers seem to employ more images, potentially creating a positive public image and diverting attention from less favorable disclosure.

Moreover, companies disclosing more extensively on CSR topics (indicated by a higher GRI Reporting Score) use fewer images. This suggests that more transparent firms that provide extensive, comparable (quantitative) information might rely less on visuals to communicate their narrative. Conversely, firms with less substantive reporting might employ visuals more strategically to enhance perceived transparency, compensating for a lack of actual information being disclosed, similar to the findings of García-Sánchez and Araújo-Bernardo (2020). These observations align with my discussion about the rationale of CSR reporting (Legitimacy Theory) and that companies might use framing and impression management to enhance perceived legitimacy and that visuals might be seen as a potent tool to overshadow less favorable aspects (e.g., (Boiral, 2013; Cho et al., 2009; Posner et al., 1976; Siano et al., 2017; Suchman, 1995)).

I observed the region to be another factor determining the frequency of picture usage. While companies in the EU tend to use more images over the whole period (for reasons I cannot directly derive from my discussions), I showed that the regulatory environment a company operates in might significantly affect and change companies' decisions on picture usage. My Difference in Difference (DiD) analysis, while slightly insignificant using robust standard errors (P-values: 0.056 / 0.064), provides a reasonable basis to believe that the Non-Financial Reporting Directive (NFRD) has a negative impact on the number of pictures used in CSR reports¹². This finding could be explained by previous literature and my results. Fiechter et al. (2022) showed that the NFRD increases reporting transparency, and my analysis indicated a decrease in picture usage through increased reporting transparency (represented by adherence to GRI guidelines -also a crucial factor for transparency used in Fiechter's research). Together, this might explain, in part, the observed effect. The impact of the NFRD on picture usage was more pronounced in sensitive industries. This greater effect can potentially be attributed to the increased scrutiny these companies face (Gamerschlag et al., 2011), which likely intensified with the NFRD, pushing these companies towards more precise reporting.

8.1.2. Determinants of Picture Content

In my analysis, I observe that the social dimension dominates the visual content in CSR reports, with almost half of all pictures depicting this theme being in line with further research (Invernizzi et al., 2022). This might be because images depicting people create a sense of closeness. In general,

relatedness and connections towards the company are essential to keeping stakeholder trust, reflecting companies' intentions to align themselves with social responsibilities. (Davison, 2010)

My findings indicate several factors that could influence picture category distribution across reports. Notably, the social dimension appears quite stable, with only the closeness of operations to customers ("Customer_Proximity") and ESG score showing a relationship with the use of social images, both being positive. It seems that companies closer to their customers might use visuals to emphasize their community involvement and social contributions. The positive association with ESG score is not straightforward. Still, it might suggest that companies with better ESG performance could be using more social images, possibly due to the rather generic content of these images and a reduced need to distinguish themselves through specialized visuals.

Interestingly, companies identified as socially sensitive do not significantly influence the use of social-themed pictures as initially expected. However, these companies show a reduced use of environmental pictures, which indicates a shift in visual focus. Although I can't definitively conclude that these companies are increasing their use of social images, the coefficients for Corporate Excellence and Social are both positive. Social is significantly higher, hinting that these companies could favor social imagery more.

Environmental images are more frequently used by companies actively engaged in environmental initiatives. This finding aligns with the dual capacity idea of visuals (Davison, 2015) and, in this case, serves a representational purpose, in contrast to a constructive purpose that I have observed so far.

Companies with a higher CO₂ intensity of revenue ("log(CO₂_Emissions_Rev)") tend to use more environmental pictures and fewer Corporate Excellence ones, possibly as a strategic move to portray themselves as environmentally responsible despite their larger environmental impact.

I observe contrasting patterns for companies sensitive to social and environmental issues (E_X_S) and those only environmentally sensitive (ENV). Environmentally sensitive firms seem to utilize more Corporate Excellence images, perhaps indicating their capability to address environmental challenges. This choice might be made because certain industries cannot clearly show direct environmental benefits due to their operational nature; instead, they demonstrate competency in managing such challenges and addressing inherent industry issues. Explaining the effect of companies being sensitive in both dimensions (Using more environmental and less Corporate Excellence pictures) is more complex, as theoretically, the same line of reasoning could apply to them. Inspecting the types of companies in my sample that fall into this category, I see that these operate in basic materials (e.g., mining and metal fabrication). Again, given the inherent limitations of their industries, these companies face structural constraints that prevent substantial changes through innovation alone. The processes they rely on are largely fixed, and meaningful improvements in their environmental impact often rely on broader advancements in

¹² While the treatment effect would be expected either after 2014 (the year the NFRD was passed) or 2016 (the reports for 2017 were the first to fall under the directive) I found the effect after 2018. While there is no definite explanation for this, I assume that this might be due to gradual adaptation effects and feedback adjustments.

high-density green energy, which is beyond their direct control. In this context, portraying themselves as pioneers of environmental innovation may not be seen as authentic due to the nature of their operations. Consequently, these firms might opt for generic environmental images in their CSR reports, possibly to create a perception of environmental awareness. This strategy could be seen as attempting to align with societal expectations and convey a commitment to environmental goals, even when operational changes are limited.

Again, contrary to what I expected, I see no correlation between ESG performance (or the different single scores) and the type of pictures used.

8.2. Limitations

Several factors reduce the generalizability of the results: the sampled data, the (classification) methodology, and the research design (in addition to no previous similar study). Attention must be paid to these limitations as they are crucial when classifying the results obtained. In the following, I will discuss these limitations and provide suggestions on how future research might deal with them.

8.2.1. Gathered Sample

In creating my sample, I introduced a selection bias by only considering companies that produce separate CSR reports. The decision to publish a separate report could have already affected the reporting behaviors I am studying through regression analyses and might, in turn, be influenced by some of the characteristics or variables under consideration in my study. Further, the direct matching approach might not lead to optimal sample groups since it only matches industry and revenue; other important factors could have been ignored. Moreover, the composition of these groups changes every year, especially in the years before 2016, when fewer companies either did not report at all or issued integrated reports. In addition, my secondary data has missing observations, which further limits the sample size for some regressions. For future studies, including integrated reports from a larger initial sample and using PSM for group creation could be beneficial. Ensuring a consistent dataset that includes the same companies each year would also be an advantage if the sample size is sufficient.

8.2.2. Picture Classification

My unique approach to content analysis allowed for the analysis of a significant number of pictures with comparable low effort, but it has limitations. I cannot apply typical error tests without manually classifying pictures and thus lack a benchmark for "correct" classification. This leads to uncertainty regarding the accuracy of the formed clusters. The technical evaluations, like the Silhouette Score, indicated rather low values (0.162), with the score ranging from -1 to 1, and 1 being the best score, suggesting that clustering was technically weak. While it does not necessarily mean the classification was ineffective, as it is reasonable that, e.g., a

picture depicting an office scene would be quite similar to pictures depicting a corporate professional, the process could, nevertheless, benefit from a deeper analysis of the words that most strongly influence the embeddings and cluster formations. First, inspecting a subset of the pictures and identifying whether the current importance of single labels accurately captures the differences that are of interest, or if other labels are better suited. Second, adjusting the weights of single labels to meet the requirements could lead to more accurate clusters. This could replace the manual adjustments I have done after the initial clustering, reducing the possibility of manual errors. Additionally, I only considered picture content without addressing attributes like color and size or the depiction of emotions (e.g., García-Sánchez and Araújo-Bernardo, 2020; Invernizzi et al., 2022), which could enhance category differentiation when added as another layer. As the Google Vision API can also analyze a picture's color composition, this could be integrated with an automated approach.

8.2.3. Research Design

My regressions could only capture a small portion of the dataset variability, signaled by low values of R-squared (< 0.2). While this may be due to intrinsic variability in the outcome variables, it could also arise from excluding some important variables not considered during the analysis due to limited research in this area. The backward elimination method also has limitations. Although it offers a systematic way of selecting variables for the model, it might have missed out on some combinations that could explain outcomes more effectively. Future studies are recommended to delve into different sets of variables along with their interactions to either confirm the persistence of the observed effects or reveal new explanatory ones. Moreover, I used only a normal linear model without considering time and unit fixed effects when running regressions on the number of pictures and picture content. This could be included in a panel regression in future studies, ideally with more evenly distributed observations over the years.

9. Conclusion and Indication for Future Research

Only a few studies have examined the use of photographs in CSR reports so far; with my study, I contribute to this field by exploring determinants regarding the amount and type of pictures used. I have done so by using a semi-automated classification approach, through which I could utilize a comparably large set of pictures and regression models.

In the paper, I analyzed why images are employed in CSR Reports, the different purposes they can serve, and what factors might influence the decision on how many and what type of pictures are used. The central theme of the research question was whether visuals are used strategically to construct a positive image of a company, potentially independent of its actual performance.

I found that industry sensitivity and measurable indicators, such as CO₂ emissions per revenue, affect the amount

and content of pictures used in CSR reports. These findings suggest that images are constructive elements in impression management, overshadowing or mitigating adverse impacts and industry flaws. However, in some cases, they could also serve as descriptive elements, e.g., when companies engaged in environmental initiatives tend to use more environmental pictures.

Further, I also assessed the impact of the Non-Financial Reporting Directive (NFRD) on using visuals in CSR reports. I discovered that the NFRD has a moderating effect, reducing the number of pictures used, which might enhance the transparency of these reports and potentially lead them to focus more on the actual disclosure of relevant information.

Given my findings, I suggest that Stakeholders must be conscious of the possibility that companies strategically use images in CSR reports. Employed as impression management techniques, they typically show an idealized image of the organization. Understanding this aspect is crucial, as the influence of visuals on perception is often subconscious, and being aware of it can assist in mitigating these subconscious effects.

Further research in this area could address the limitations of this study and test whether these results hold for different samples or explore different variables and combinations that may affect the use of images. It would also be beneficial to examine the impact of the NFRD on the content of visuals in CSR reports. Furthermore, as in this paper, I only investigated the patterns influencing the usage of images, it would also be interesting to study whether and to what extent the amount and content of pictures influence the valuation of a company's ESG performance and perception of sustainability efforts.

References

- Accenture. (2022). Unlocking the Global Pathways to Resilience, Growth, and Sustainability for 2030: The 12th United Nations Global Compact-Accenture CEO Study, 2022. <https://www.accenture.com/content/dam/accenture/final/accenture-com/document/Accenture-CEO-Study-United-Nations-Global-Compact.pdf>
- Baron, D. P. (2001). Private Politics, Corporate Social Responsibility, and Integrated Strategy. *Journal of Economics & Management Strategy*, 10(1), 7–45. <https://doi.org/10.1111/j.1430-9134.2001.00007.x>
- Benabou, R., & Tirole, J. (2010). Individual and Corporate Social Responsibility. *Economica*, 77(305), 1–19. <https://doi.org/10.1111/j.1468-0335.2009.00843.x>
- BlackRock. (2020). Sustainability goes mainstream, 2020. <https://img.lair.co/cms/2021/05/28202727/blackrock-sustainability-survey.pdf>
- Boiral, O. O., & Henri, J.-F. (2017). Is Sustainability Performance Comparable? A Study of GRI Reports of Mining Organizations. *Business & Society*, 56(2), 283–317. <https://doi.org/10.1177/0007650315576134>
- Boiral, O. O. (2013). Sustainability reports as simulacra? A counter-account of A and A+ GRI reports. *Accounting, Auditing & Accountability Journal*, 26(7), 1036–1071. <https://doi.org/10.1108/AAAJ-04-2012-00998>
- Bonetti, P., Cho, C. H., & Michelon, G. (2023). Environmental Disclosure and the Cost of Capital: Evidence from the Fukushima Nuclear Disaster. *European Accounting Review*, 1–29. <https://doi.org/10.1080/09638180.2023.2203410>
- Brammer, S., & Millington, A. (2005). Corporate Reputation and Philanthropy: An Empirical Analysis. *Journal of Business Ethics*, 61(1), 29–44. <https://doi.org/10.1007/s10551-005-7443-4>
- Branco, M. C., & Rodrigues, L. L. (2008). Factors Influencing Social Responsibility Disclosure by Portuguese Companies. *Journal of Business Ethics*, 83(4), 685–701. <https://doi.org/10.1007/s10551-007-9658-z>
- Business Roundtable. (2019). Business Roundtable Redefines the Purpose of a Corporation to Promote 'An Economy That Serves All Americans'. <https://www.businessroundtable.org/business-roundtable-redefines-the-purpose-of-a-corporation-to-promote-an-economy-that-serves-all-americans>
- Byrd, J., Hickman, K., Baker, C., & Cohanier, B. (2017). Corporate Social Responsibility Reporting in Controversial Industries. <https://doi.org/10.2139/ssrn.2894789>
- Carleton, T. A., & Hsiang, S. M. (2016). Social and economic impacts of climate. *Science*, 353(6304). <https://doi.org/10.1126/science.aad9837>
- Cho, C., Phillips, J., Hageman, A., & Patten, D. (2009). Media richness, user trust, and perceptions of corporate social responsibility: An experimental investigation of visual web site disclosures. *Accounting, Auditing and Accountability Journal*, 22(6), 933–952. <https://doi.org/10.1108/09513570910980481>
- Chong, S., Narayan, A. K., & Ali, I. (2019). Photographs depicting CSR: captured reality or creative illusion? *Pacific Accounting Review*, 31(3), 313–335. <https://doi.org/10.1108/PAR-10-2017-0086>
- Christensen, H. B., Hail, L., & Leuz, C. (2021). Mandatory CSR and sustainability reporting: economic analysis and literature review. *Review of Accounting Studies*, 26, 1176–1248. <https://doi.org/10.1007/s11142-021-09609-5>
- Davison, J. (2008). Rhetoric, repetition, reporting and the “dot.com” era: words, pictures, intangibles. *Accounting, Auditing & Accountability Journal*, 21(6), 791–826. <https://doi.org/10.1108/09513570810893254>
- Davison, J. (2010). [In]visible [in]tangibles: Visual portraits of the business elite. *Accounting, Organizations and Society*, 35(2), 165–183. <https://doi.org/10.1016/j.aos.2009.03.003>
- Davison, J. (2015). Visualising accounting: an interdisciplinary review and synthesis. *Accounting and Business Research*, 45(2), 121–165. <https://doi.org/10.1080/00014788.2014.987203>
- Davison, J., & Warren, S. (2009). Imag[in]ing accounting and accountability. *Accounting, Auditing & Accountability Journal*, 22(6), 845–857. <https://doi.org/10.1108/09513570910980436>
- Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting & Economics (JAE)*, Forthcoming. <https://doi.org/10.2139/ssrn.2741682>
- Erickson, S. L., Weber, M., & Segovia, J. (2011). Using Communication Theory to Analyze Corporate Reporting Strategies. *Journal of Business Communication*, 48(2), 207–223. <https://doi.org/10.1177/0021943611399728>
- Fiechter, P., Hitz, J.-M., & Lehmann, N. (2022). Real Effects of a Widespread CSR Reporting Mandate: Evidence from the European Union's CSR Directive. *Journal of Accounting Research*, 60(4), 1499–1549. <https://doi.org/10.1111/1475-679X.12424>
- Gamerschlag, R., Möller, K., & Verbeeten, F. (2011). Determinants of voluntary CSR disclosure: empirical evidence from Germany. *Review of Managerial Science*, 5, 233–262. <https://doi.org/10.1007/s11846-010-0052-3>
- García-Sánchez, I.-M., & Araújo-Bernardo, C.-A. (2020). What colour is the corporate social responsibility report? Structural visual rhetoric, impression management strategies, and stakeholder engagement. *Corporate Social Responsibility and Environmental Management*, 27(2), 1117–1142. <https://doi.org/10.1002/csr.1869>
- Gatti, L., Pizzetti, M., & Seele, P. (2021). Green lies and their effect on intention to invest. *Journal of Business Research*, 228–240. <https://doi.org/10.1016/j.jbusres.2021.01.028>
- Global Reporting Initiative. (2006). Sustainability reporting guidelines. Amsterdam, The Netherlands.
- Google. (2024). Cloud Vision documentation. <https://cloud.google.com/vision/docs?hl=de>

- Grougiou, V., Dedoulis, E., & Leventis, S. (2016). Corporate Social Responsibility Reporting and Organizational Stigma: The Case of “Sin” Industries. *Journal of Business Research*, 69(2), 905–914. <https://doi.org/10.1016/j.jbusres.2015.06.041>
- Hahn, R., & Kühnen, M. (2013). Determinants of sustainability reporting: a review of results, trends, theory, and opportunities in an expanding field of research. *Journal of Cleaner Production*, 59, 5–21. <https://doi.org/10.1016/j.jclepro.2013.07.005>
- Harper, D. (2002). Talking about pictures: A case for photo elicitation. *Visual Studies*, 17(1), 13–26. <https://doi.org/10.1080/14725860220137345>
- He, C., & Loftus, J. (2014). Does environmental reporting reflect environmental performance? *Pacific Accounting Review*, 26(1/2), 134–154. <https://doi.org/10.1108/PAR-07-2013-0073>
- Hrasky, S. (2012). Visual disclosure strategies adopted by more and less sustainability-driven companies. *Accounting Forum*, 36(3), 154–165. <https://doi.org/10.1016/j.accfor.2012.02.001>
- Huang, C.-L., & Kung, F.-H. (2010). Drivers of Environmental Disclosure and Stakeholder Expectation: Evidence from Taiwan. *Journal of Business Ethics*, 96, 435–451. <https://doi.org/10.1007/s10551-010-0476-3>
- Humaira, H., & Rasyidah, R. (2018). Determining The Appropriate Cluster Number Using Elbow Method for K-Means Algorithm. <https://doi.org/10.4108/eai.24-1-2018.2292388>
- Invernizzi, A. C., Bellucci, M., Acuti, D., & Manetti, G. (2022). Form and substance: Visual content in CSR reports and investors’ perceptions. *Psychology & Marketing*, 39(5), 974–989. <https://doi.org/10.1002/mar.21635>
- Kitzmueller, M., & Shimshack, J. (2012). Economic Perspectives on Corporate Social Responsibility. *Journal of Economic Literature*, 50(1), 51–84. <https://doi.org/10.1257/jel.50.1.51>
- Legendre, T. S., Jarvis, N., Kang, Y., Jamal, G., & Jackson, J. (2020). Rescuing imperfect produce: The effects of stigma disclosure strategies, controllability, and aesthetics. *International Journal of Hospitality Management*, 85, 102443. <https://doi.org/10.1016/j.ijhm.2019.102443>
- Nielsen, A. E., & Thomsen, C. (2018). Reviewing corporate social responsibility communication: a legitimacy perspective. *Corporate Communications: An International Journal*, 23(4), 492–511. <https://doi.org/10.1108/CCLJ-04-2018-0042>
- Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, 83(2), 157–171. <https://doi.org/10.1037/0033-295X.83.2.157>
- Preston, A. M., Wright, C., & Young, J. J. (1996). IMAG[IN]ING ANNUAL REPORTS. *Accounting, Organizations and Society*, 21(1), 113–137. [https://doi.org/10.1016/0361-3682\(95\)00032-5](https://doi.org/10.1016/0361-3682(95)00032-5)
- Rämö, H. (2011). Visualizing the Phronetic Organization: The Case of Photographs in CSR Reports. *Journal of Business Ethics*, 104, 371–387.
- Riveros-Gavilanes, J. M. (2023). Testing Parallel Trends in Differences-in-Differences and Event Study Designs: A Research Approach Based on Pre-Treatment Period Significance. *Journal of Research, Innovation and Technologies (JoRIT)*, 2(4), 226–237. [https://doi.org/10.57017/jorit.v2.2\(4\).07](https://doi.org/10.57017/jorit.v2.2(4).07)
- Serafeim, G., & Amel-Zadeh, A. (2017). Why and How Investors Use ESG Information- Evidence from a Global Survey. *Financial Analysts Journal*, 74(3), 87–103. <https://doi.org/10.2139/ssrn.2925310>
- Siano, A., Vollero, A., Conte, F., & Amabile, S. (2017). “More than words”: Expanding the taxonomy of greenwashing after the Volkswagen scandal. *Journal of Business Research*, 71, 27–37. <https://doi.org/10.1016/j.jbusres.2016.11.002>
- Song, K., Tan, X., Qin, T., Lu, J., & Liu, T.-Y. (2020). MPNet: Masked and Permuted Pre-training for Language Understanding. <http://arxiv.org/pdf/2004.09297v2>
- Stolowy, H., & Paugam, L. (2018). The expansion of non-financial reporting: an exploratory study. *Accounting and Business Research*, 48(5), 525–548. <https://doi.org/10.1080/00014788.2018.1470141>
- Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *The Academy of Management Review*, 20(3), 571–610. <https://doi.org/10.2307/258788>
- Tate, W. L., Ellram, L. M., & Kirchoff, J. L. (2009). CORPORATE SOCIAL RESPONSIBILITY REPORTS: A THEMATIC ANALYSIS RELATED TO SUPPLYCHAIN MANAGEMENT. *Journal of Supply Chain Management*, 46(1), 19–44. <https://doi.org/10.1111/j.1745-493X.2009.03184.x>
- Townsend, C., & Shu, S. B. (2010). When and how aesthetics influences financial decisions. *Journal of Consumer Psychology*, 20(4), 452–458. <https://doi.org/10.1016/j.jcps.2010.06.013>
- Tversky, B. (1974). Eye fixations in prediction of recognition and recall. *Memory & Cognition*, 2, 275–278. <https://doi.org/10.3758/BF03208995>
- Vasue, E. S., & Howe, A. C. (1989). The effect of visual and verbal modes of presentation on children’s retention of images and words. *J. Res. Sci. Teach.*, 26(5), 401–407. <https://doi.org/10.1002/tea.3660260505>
- Wang, Y. (2019). Single Training Dimension Selection for Word Embedding with PCA. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3597–3602. <https://doi.org/10.18653/v1/D19-1369>
- Worrell, E., Bernstein, L., Roy, J., Price, L., & Harnisch, J. (2009). Industrial energy efficiency and climate change mitigation. *Energy Efficiency*, 2, 109–123. <https://doi.org/10.1007/s12053-008-9032-8>
- Yilmaz, Ö., Achenie, L. E. K., & Srivastava, R. (2007). Systematic tuning of parameters in support vector clustering. *Mathematical Biosciences*, 205, 252–270. <https://doi.org/10.1016/j.mbs.2006.09.013>



Determinants of Corporate Bond Mutual Fund Flows

Vanessa Jeske

University of Stuttgart

Abstract

This paper examines the determinants of investor flows into U.S. corporate bond mutual funds, with a focus on monetary policy and fund-specific characteristics during the COVID-19 crisis. These funds, as non-bank financial intermediaries, are vulnerable to sudden investor redemptions due to liquidity mismatches. Using monthly data from 2001 to 2021, the analysis applies panel regressions with fund style and time fixed effects to assess how monetary policy, fund characteristics, and market conditions influence investor behavior. Results show that higher effective federal funds rates are significantly associated with reduced fund flows. Past flows and performance rankings are strong predictors of current flows, while fund cash holdings matter mainly in riskier fund types. During the COVID-19 crisis, flow sensitivity to interest rate changes intensified. Although Federal Reserve policy announcements in spring 2020 coincided with a quick return of inflows, the findings emphasize ongoing structural fragility. By analyzing flow dynamics alongside macroeconomic factors and policy responses, this research contributes to understanding the determinants of corporate bond mutual fund flows and the complex role of central bank actions during periods of systemic stress.

Keywords: corporate bond mutual funds; COVID-19 crisis; federal reserve policy; fund flows; liquidity risk

1. Introduction

US bond mutual funds have expanded significantly over the last decade and gained importance. By 2021, they captured \$5.6 trillion of the US economy, i.e., 10% of the US bond market, and received \$2.6 trillion in net new cash flow (see Figure 1).

Corporate bond mutual funds (CBMF) make up a substantial portion of that bond market. Corporate bonds themselves are a pivotal tool for corporate debt financing yet tend to be highly illiquid.¹ Since CBMFs are not associated with banks, they are classified as non-bank financial intermediaries (NBFIs) which are less strictly regulated than banks and

have a crucial macroeconomic role. Indeed, in 2021 the Global Monitoring Report on Non-Bank Financial Intermediation documented that NBFIs captured \$14.7 trillion (62% of the US GDP) and, worryingly, considered them vulnerable to runs.² This run risk is based on a fundamental liquidity mismatch as most CBMFs offer daily withdrawal schedules and therefore continuously engage in major liquidity transformations. The liquidation costs are internalized by the fund, encouraging a first-mover effect. Consequently, CBMFs raise concerns regarding their systematic financial fragility and market resilience.

The COVID-19 crisis provided a global stress test as it led to a corporate bond liquidity crisis in March 2020. Adam Lollo from Citigroup Inc. reflected: “The 2008 financial crisis was a car crash in slow motion,” ... “This was like, ‘Boom!’”³ The corporate bond market experienced extreme

First, I would like to sincerely thank Prof. Dr. Philipp Schuster for his constructive guidance and academic support throughout my thesis. I am especially grateful to my supervisor, Franziska Weishaupt, M.Sc., for her insightful feedback, generous availability, as well as her support in the data collection process. Finally, I would like to thank my mother, Juliet Jeske, for her unwavering support and Charis Lieberum and Theresa Thölking for their enthusiastic encouragement throughout this journey.

¹ Bao et al. (2011), p. 911 f.

² cp. Financial Stability Board (2024a), URL see References

³ cp. Baer (2020), URL see References

⁴ cp. Investment Company Institute (2022), p. 59.



Figure 1: US bond mutual funds net inflows⁴

liquidity strains and transaction costs.⁵ Combined with extreme selling pressures, the Bank of America considered the bond market “ ‘basically broken’ ”.⁶ The economic situation remained dire until the Federal Reserve System (Fed) committed \$75 billion of equity for corporate credit facilities aimed at supporting the bond market.⁷ The medium-term effects of the Fed’s policy intervention require further study.

The existing literature on CBMF flows is far less comprehensive than similar research on equity funds. This research gap is slowly being bridged; however, authors primarily focus on the isolated influence of fund-specific or macro factors on the performance by way of the flow-performance relationship. Additionally, few macro determinants have been extensively studied. Fund flow dynamics themselves have not received much focused attention with their systematic liquidity risk being potentially neglected. Conversely, researchers comprehensively investigated the COVID-19 corporate bond liquidity crisis, providing ample insights into bond evolution during the immediate pandemic. Since relevant studies were chiefly published in 2020 and 2021, they could only analyze short-term effects. Research on the medium-term implications becomes possible as time passes, but literature with data sample periods up to the year-end of 2021 or beyond is still few and far between. Currently, few papers integrate fund characteristics, macro conditions, monetary policy, and pandemic crisis in one cohesive study.

This thesis aims to fill this gap and asks: First, what are the determinants of corporate bond mutual fund flows, with a particular focus on the effective federal funds rate, effective federal funds rate change, and cash ratio; second, how

do fund styles affect these determinants; and third, how did the COVID-19 crisis influence CBMF flows in the medium term? First, I investigate the influence of fund-specific and macro factors on fund flows in a systematic fund flows analysis spanning 20 years in order to explore, confirm or reject well-studied correlations. Then, I explore if the impact of these determinants varies within fund styles, e.g., high-yield vs. corporate vs. nontraditional bond mutual funds. Finally, I want to study the short-term and, notably, the medium-term impact of the COVID-19 pandemic on fund flows using data extending up to December 2021. Since the pandemic was an exogenous stress event, I focus the pandemic dedicated case study on the influence of the macro factors.

The thesis continues in chapter 2 with an explanation of the underlying theory. Firstly, subchapter 2.1 presents the fundamentals of CBMFs. Secondly, subchapter 2.2 establishes the theoretical framework of fund flows and the fragility of financial systems. Lastly, subchapter 2.3 introduces the macro perspective and chronicles the corporate bond market liquidity crisis. Chapter 3 reviews the current state of research, examining the current relevant literature on corporate bond mutual fund flows and how they were impacted by the pandemic. Chapter 4 describes the data and methodology used, namely the construction of the CBMF data sample, summary statistics of the fund information and macro data and the preparation and calculation of the independent and dependent variables. Chapter 5 presents the core analysis and results of the thesis. Subchapter 5.1 begins with an examination of systematic fund flow dynamics through Ordinary Least Squares (OLS) and Fixed Effects (FE) regressions. Subchapter 5.2 delves deeper into the systematic dynamics by conducting a split-sample analysis based on fund styles. Finally, subchapter 5.3 offers a case study of the COVID-19 pandemic’s impact on CBMFs. Chapter 6

⁵ cf. O’Hara and Zhou (2021), p. 46 f.

⁶ cp. Idzelis (2020), URL see References

⁷ cp. Board of Governors of the Federal Reserve System (2020c), URL see References

concludes the thesis with a summary and brief discussion of future considerations.

2. Theory

This chapter explores the theoretical foundations surrounding US CBMFs and their role in the US financial market. An overview of the fundamentals is provided. It dives deeper into the theory behind fund flow dynamics and financial fragility and proceeds with illuminating the influence of the macro condition by means of macro factors and central bank policy on CBMFs. Finally, the state of the corporate bond market during the COVID-19 crisis is summarized.

2.1. Corporate Bond Mutual Funds: Fundamentals

2.1.1. Essentials, Risks and Fund Styles

A corporate bond is a debt security issued by a company to raise funding for business operations, capacity, or investments.⁸ Investors can bundle their funds in professionally managed mutual funds which use the aggregated capital to purchase diversified bond portfolios. Most mutual funds are open-end, i.e., they can issue unlimited shares and grant their investors daily withdrawal rights. Many funds have a distinct investment profile and fund style, which allows investors to select funds matching their own requirements.

Corporate bonds can be vulnerable to default, interest rate, economic, liquidity and inflation risks. Default risk or credit risk describes the probability of the bond issuer defaulting, e.g., through insolvency, while liquidity risk describes the potential difficulty bondholders might have in selling and converting the bond value to cash at will.⁹ Interest rate risk occurs when market interest rates rise, new bonds offer better terms to investors, making the existing bond prices lose value; inflation risk describes the danger that rising inflation reduces the effective value of the bond coupons and lowers purchasing power of consumers, potentially lowering the profits of bond issuers. Finally, the economic risk is the risk of economic turmoil inducing investors to shift to assets with higher credit ratings and sell a majority of their bond holdings.¹⁰ These risks, which influence the capital allocations of investors and thereby fund flows, differ between fund categories.

Over time, many different CBMF fund styles have emerged to suit investors performance, stability, and risk needs. To permit a more detailed analysis of the influence of the fund style on the fund flows, we will first be differentiating between the nine different categories of CBMFs as assigned by Morningstar. US high-yield (HY) bond funds invest in “lower-quality bonds” and thus “generally offer higher yields [...] but they are also more vulnerable to economic and

credit risk”.¹¹ US short-term bond funds “invest primarily in corporate and other investment-grade U.S. fixed-income issues and have durations of one to 3.5 years” and are “less sensitive to interest rates,” which satisfies more risk-averse investors.¹² In sum, short-term bond funds have a low interest rate and credit risk compared to other fund styles and, due to the short maturation duration, little liquidity risk.

US intermediate core Bond Funds mostly consist of “investment-grade U.S. fixed-income issues . . . and typically hold less than 5% in below-investment-grade exposures” with an intermediate “effective duration of the Morningstar Core Bond Index”, which are “a measure of interest-rate sensitivity”.¹³ Similarly, US intermediate core-plus bond funds also consist of intermediate investment-grade (IG) bonds but have “greater flexibility”.¹⁴

US corporate bond funds are composed of “investment-grade bonds issued by corporations in US dollars, which tend to have more credit risk than government or agency-backed bonds”.¹⁵ Additionally, corporate bonds may have a significant inflation and economic risk as the issuing corporations’ successes rely on the economic health of their consumers and their purchasing power. In a global or sector-wide economic downturn, the economic risk would be significant as investors globally try to shift to less risky assets. Of course, investors of this category may be motivated by non-economic factors, notably a belief in a corporation or sector, to incorporate this bond type in their portfolio. US multisector bond funds have diversified their investments across a wide range of domestic and foreign fixed-income sectors.¹⁶ Due to their larger credit risk, they tend to offer higher total returns but suffer significant losses during equity market stress periods.¹⁷ US long-term bond funds primarily hold “corporate and other investment-grade U.S. fixed-income issues” and have durations of six years or more, making them prone to interest rate risk.¹⁸

US nontraditional bond funds consist of funds with disparate strategies and flexible mandates, e.g., absolute return portfolios which aim to produce returns independently from the bond market, unconstrained portfolios which boast very high flexibility and allocation size and minimum volatility portfolios which seek to reduce volatility despite high credit risk and potentially high return investments.¹⁹ Their goal is to outperform specific benchmarks or focus on other benchmarks. Finally, US bank loan funds invest in “floating-rate bank loans and other floating-rate securities” and “in exchange for their credit risk . . . , these loans offer higher interest rates that typically float above a common short-term

¹¹ cf. Morningstar (2025b), URL see References

¹² cf. Morningstar Office (2023e), URL see References

¹³ cp. Bush (2019b), URL see References

¹⁴ cp. Bush (2019b), URL see References

¹⁵ cf. Morningstar (2025a), URL see References

¹⁶ cf. Morningstar Office (2023c), URL see References

¹⁷ cf. Bush (2019a), URL see References

¹⁸ cp. Morningstar Office (2023b), URL see References

¹⁹ cp. Morningstar Office (2023d), URL see References

⁸ cf. U.S. Securities and Exchange Commission (2013a), URL see References

⁹ cf. U.S. Securities and Exchange Commission (2013b), URL see References

¹⁰ cf. U.S. Securities and Exchange Commission (2013a), URL see References

benchmark".²⁰ This synopsis of the different fund categories forms the conceptual foundation of the Morningstar Category split sample analysis in subchapter 5.2 and allows for a substantive interpretation of the fund style as an independent variable.

2.1.2. Financial Intermediaries and Investors

CBMFs invest primarily in corporations. After bonds have been issued and sold on the primary market, investors use over-the-counter (OTC) brokerages to purchase and trade bonds and mutual fund shares.²¹ When an investor wants to redeem their investment, they receive the total return after fees are deducted. These OTC dealers are classified as NBFIs with the potential to carry systematic risks if engaged in liquidity transformations.²² Liquidity transformation in mutual funds describes the process of converting non-liquid assets like bonds and stocks to cash, mostly to satisfy daily investor withdrawals.

The Global Monitoring Report on Non-Bank Financial considers fixed-income funds to be vulnerable to runs.²³ This is because the liquidation costs accompanying investor redemptions are not paid by the exiting investors but instead internalized by the fund, which reduces the return of the remaining investors and creates a first-mover incentive.²⁴ Thus, investors who expect share sales are motivated to sell first to become advantageous first movers, which turns large redemptions into a self-fulfilling prophecy. Corporate bond funds especially suffer from a perpetual liquidity mismatch because their main asset category, corporate bonds, are generally illiquid: They are traded much more rarely than stocks, can have substantial transaction costs and have a higher credit risk than e.g., government bonds.²⁵ To combat these risks, mutual funds are likely to have a high level of cash on hand, a so-called cash buffer. Nevertheless, such measures are only effective if they satisfy the fund investors.

CBMF fund investors are typically rather conservative, desiring low risk and a higher predictability, thus being satisfied with moderate returns. With the advent of better and more accessible technology, the number of financial intermediaries and retail investors increased.²⁶ Consequently, their behavior and psychology gained prominence in analyzing and predicting fund flow movements. Human behavior is known to diverge from the ideal financial agent, *homo economicus*, following detrimental behavioral patterns such as overreactions, heuristic simplification, and loss aversion (which often translates to risk aversion).²⁷ In times of economic uncertainty and distress, investors often engage in flight-to-liquidity and flight-to-quality behaviors. Flight-to-liquidity is the capital allocation towards assets with higher

liquidity like US Treasury bonds and cash.²⁸ Meanwhile, flight-to-quality or flight-to-safety refers to investors moving their capital from higher-risk to lower-risk assets.²⁹ Reverse flight-to-quality occurs when investors sell higher-rated assets first to acquire liquidity or because they are unable to sell lower-rated assets.³⁰ These behavioral phenomena shape fund flows and often align with the distinctive fund styles. To understand the nuances of these dynamics, it is necessary to explore the diverse fund categories. Next, the dependent variable, i.e., the fund flow, will be considered.

2.2. Fund Flows: Theoretical Framework and Financial Fragility

The fund flow needs to be calculated in accordance with the literature standard to be comparable with other published fund flow studies. Thus, I follow the calculation method of Sirri and Tufano and Chevalier and Ellison.³¹

$$Flows_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1} * (1 + r_{i,t})}{TNA_{i,t-1}} \quad (1)$$

The current flows of fund *i*, $Flows_{i,t}$, are calculated as the difference between the current total net assets, $TNA_{i,t}$, and the prior total net assets, $TNA_{i,t-1}$, multiplied by the returns relative to the current total net returns, $r_{i,t}$. Investor flow dynamics are known to be affected by a variety of factors, the most prominent of which are persistence and performance chasing. Persistence or momentum describes the phenomenon of past flows predicting current flows, i.e., funds with past inflows receiving more inflows.³² This leads to a highly significant and strong correlation between the recent and current flows. Next, investors who chase performance value the best return performance and continuously reallocate their capital to the best performing funds in their pursuit. These winner funds receive disproportionately large inflows, and accordingly, loser funds experience disproportionately large outflows. For this reason, the return performance, proxied by the fund alpha or performance rank, has a highly significant and large effect on fund flows. Furthermore, fund flows are known to be driven by fund size, fund age, and return volatility, while the influence of the expense ratio is reportedly negative.³³ Another key characteristic of a given fund is its flow-performance relationship. CBMFs have been shown to have a more concave flow-performance relationship which suggests that bond investors are particularly responsive to poor fund performance and penalize it with disproportionate outflows but are not as sensitive to superior

²⁰ cp. Morningstar Office (2023a), URL see References

²¹ cf. Kramer (2024), URL see References

²² cf. Financial Stability Board (2024b), URL see References

²³ cp. Financial Stability Board (2024a), URL see References

²⁴ cp. Q. Chen et al. (2010), p. 597

²⁵ cf. Q. Chen et al. (2010), p. 597

²⁶ cf. PIMCO (n.d.), URL see References

²⁷ cf. Hirshleifer (2015), p. 137–144

²⁸ cf. Longstaff (2004), p. 512 f.

²⁹ cf. Beber et al. (2009), p. 925

³⁰ cf. Ma et al. (2022), p. 4674 f.

³¹ cf. Chevalier and Ellison (1997), p. 1173 and cp. Sirri and Tufano (1998), p. 1594

³² cf. Sirri and Tufano (1998), p. 1619

³³ cf. Sirri and Tufano (1998), p. 1599, cf. Q. Chen et al. (2010), p. 406, cf. Goldstein et al. (2017), 602

performance. This leads to an inherent instability in the bond market which raises the question of financial fragility.

Financial fragility is defined as the propensity of a system to falter due to small liquidity demand shocks leading to disproportionately large disruptions, e.g., high asset-price volatility and bank defaults.³⁴ More generally, it describes the susceptibility and vulnerability of a financial system or market to a financial crisis.³⁵ The bigger the system, the bigger the potential emergency. Complementarily, financial stability or resilience describes the ability to absorb and weather these shocks without outsize impact. As NBFIs, the growing corporate bond debt markets and mutual funds constitute a key component of the bond financial system and, due to its inherent liquidity mismatch, liquidity shocks would be especially dangerous. This raises concerns about their financial fragility as a liquidity demand shock in response to mass redemptions during a macro crisis seems inevitable. Evidently, the macro conditions play a crucial role for CBMF resilience.

2.3. Macroeconomic Determinants of Fund Flow Dynamics

2.3.1. Macroeconomic Conditions and Monetary Policy Transmission Channels

The returns of CBMFs are dependent on multiple macro factors. As corporate bonds carry a higher credit risk than Treasury bills, investors expect to be compensated with a higher return. This is operationalized as the yield spread, which is calculated as the difference between two different bonds of the same maturity but different credit ratings. Therefore, it is dependent on the performance of baseline rates of safer investments like the 4-week Treasury bill, 3-month Treasury bill for short-term bonds or the intermediate government index. Additionally, bonds carry interest rate risks and thus are impacted by the implied stock market volatility index (VIX), which measures the aggregate market liquidity. For risk-tolerant investors, the default spread is also relevant as it measures the excess return of high-yield bonds compared to the intermediate government bond index. Finally, companies often issue stocks in addition to bonds, which is why their bonds are also impacted by the performance of the stock market index. In short, any macro variable that influences the CBMF yield spread could be relevant for fund flow analyses.³⁶

As discussed, investors tend towards loss and risk aversion. Central banks can influence investor behavior through monetary policy transmission channels (MPTC), e.g., their policy decisions can affect markets indirectly by influencing interest rates, credit supply and rate expectations.³⁷ The interest rate channel theory examines the effects of monetary tightening and loosening. The credit channel theory asserts that debt financing costs caused by “informational frictions ... worsen” during illiquid periods and thereby amplify the impact of monetary policies.³⁸ Therefore, these

MPTCs influence the benefits of higher-risk assets such as CBMFs and are relevant for interpreting investor flows. As the US-American central bank and maker of national monetary policies, the Fed is a vital player in these transmission channels. The Federal Reserve Bank of New York, a supporting reserve bank, publishes the daily effective federal funds rate (FEDFUNDS).³⁹ This rate affects the lending costs of loans among domestic depository institutions like banks.⁴⁰ These borrowing costs indirectly trickle down to other liquidity costs and thus can impact debt financing instruments across the board. This matters particularly during a financial crisis like the COVID-19 liquidity crisis.

2.3.2. The Corporate Bond Market during the COVID-19 Liquidity Crisis

By early March, the COVID-19 pandemic had fully disrupted the US and global financial markets. In fixed-income markets, the extreme instability and uncertainty led to extreme yield spreads in corporate bonds as corporations struggled to operate during imposed restrictions and lockdowns (see Figure 2).⁴¹ This chain of events included an extreme bond price crash and liquidity strain in corporate bond markets.⁴²

The bond crisis peaked within three weeks as fixed-income markets inevitably became highly illiquid, and investors turned desperate.⁴³ IG bond funds suffered especially. Two primary reasons were selling pressures and lacking liquidity. Selling pressures increased immensely as mutual funds in desperate need of cash sold Treasury securities worth \$266 billion in a reverse flight-to-quality while liquidity provision faltered as OTC dealers were unable (due to full capacity) or unwilling to take bonds into inventory and supply liquidity.⁴⁴ At this point, the Fed directly entered the corporate bond market as a “market maker of last resort” and bought illiquid assets through its newly created corporate credit facilities.⁴⁵ The chronology is as follows: On March 23, the Fed announced the PMCCF and SMCCF with their initial term sheets: The PMCCF’s role was to act as a liquidity backstop for corporate debt by buying bonds directly from and providing loans to eligible issuers, while the SMCCF’s purpose was to purchase secondary market corporate bonds from eligible issuers.⁴⁶ This served as a liquidity backstop for US corporations in severe need of liquid funds to continue operating. On April 9, the Fed increased the Treasury capital to \$50 billion and \$25 billion of equity and included HY CBMFs.⁴⁷

³⁹ cf. Board of Governors of the Federal Reserve System (2025), URL see References

⁴⁰ cf. Federal Reserve Bank of New York (2025), URL see References

⁴¹ cf. O’Hara and Zhou (2023), p. 56 f.

⁴² cp. O’Hara and Zhou (2023), p. 57

⁴³ cf. O’Hara and Zhou (2023), p. 58

⁴⁴ cf. O’Hara and Zhou (2021), p. 57

⁴⁵ cf. O’Hara and Zhou (2021), p. 46

⁴⁶ cf. Board of Governors of the Federal Reserve System (2020a, 2020b, 2020d), URL see References

⁴⁷ cf. Board of Governors of the Federal Reserve System (2020c), URL see References

³⁴ cf. Allen and Gale (2004), p. 1027

³⁵ cf. Lagunoff and Schreft (2001), p. 220

³⁶ cf. Y. Chen and Qin (2017), pp. 7 f.

³⁷ cf. European Central Bank (2025), URL see References

³⁸ cf. Mishkin (1995), p. 4, cf. Bernanke and Gertler (1995), p. 35

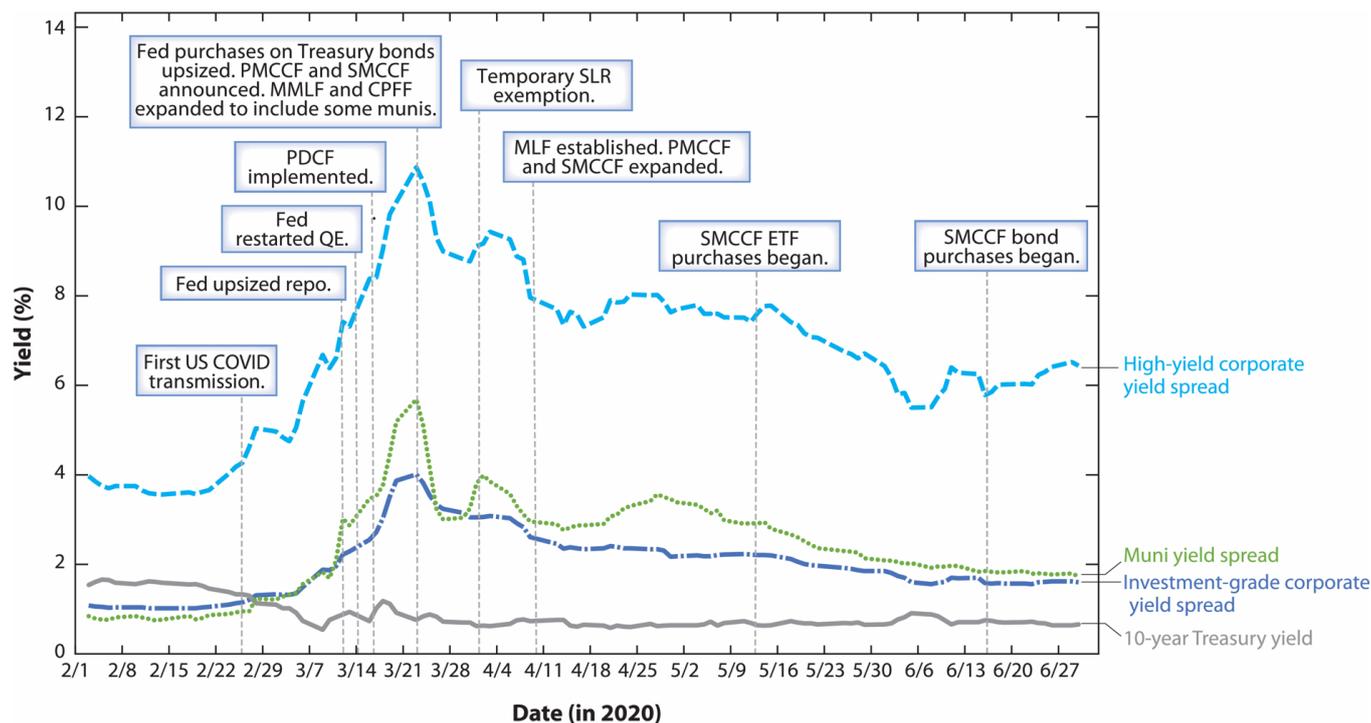


Figure 2: US COVID-19 liquidity crisis evolution and US macro policy responses⁴⁸

The PMCCF launched on June 29 and the SMCCF launched on May 12; both had ceased operations by December 13, 2020.⁴⁹ Through these actions, the Fed improved aggregate bond market liquidity. Regarding the liquidity demands placed on mutual funds, the Fed proved most effective as the SMCCF announcement reversed outflows, in particular for more fragile funds.⁵⁰ These effects continued and created “positive spillovers to primary bond markets and to other funds holding similar assets”.⁵¹

3. Literature Review

This chapter reviews the relevant, existing literature on CBMF flows, focusing on the determinants and performance relationships identified by Chen and Qin and Goldstein et al. (2017) as well as the impact of macro factors explored by Kuong et al. (2024). Additionally, it examines the effects of the COVID-19 pandemic on bond market flows, drawing insights from studies by Falato et al. (2021), Boyarchenko et al. (2022) and O’Hara and Zhou (2021). These analyses provide a comprehensive understanding of the dynamics influencing CBMFs. Lastly, it will situate the research question within existing research.

⁴⁸ cp. O’Hara and Zhou (2023), p. 57

⁴⁹ cf. Federal Reserve Bank of New York (2020), URL see References

⁵⁰ cf. Falato et al. (2021), p. 37

⁵¹ cf. Falato et al. (2021), p. 37

3.1. Systematic Fund Flow Dynamics

3.1.1. Influence of Macroeconomic Variables on Fund Flows

The macro influence on CBMF flows is a central theme across many studies. Chen and Qin study the determinants of CBMF flows and their flow-performance relation, confirming for the first time in academic literature the sensitivity of investor flows to the recent macro conditions. They find a positive association between the BOND, default spread, OPTION factor, stock market return and VIX index and the money flows, respectively. The correlation to the three-month Treasury bill rate is negative. Their results suggest that rising short-term rates reduce inflows into corporate bond funds.⁵²

Kuong et al. expand the literature on the influence of macros on CBMF flows and aggregate fund fragility by exploring the impact of changes in the Federal Funds Target rate (Δ FFTar). Firstly, they document that aggregate bond outflow (inflow) increases (decreases) correspond to FFTar raises (cuts). Secondly, they regress annual and monthly fund flows as dependent variable on FFTar changes and macro controls like the logarithm of the VIX index, default risk changes and a temporal dummy variable for the COVID-19 crisis, among others. The results show a highly significant correlation. Hence, Kuong et al. put forth the following mechanism: Market participants anticipate the decrease in net asset value when learning of planned FFTar increases before FOMC meetings and thus redeem their shares, which suffer from stale overpricing, at an increased rate both before

⁵² cf. Y. Chen and Qin (2017), pp. 1-17

and after those meeting dates. They term this phenomenon outflow- Δ FFTar sensitivity.⁵³

Goldstein et al. analyze investor flow dynamics into CBMFs and study the flow-performance relationship. They establish the escalating effects of asset and corporate bond market illiquidity, two mechanisms behind bond liquidity mismatches. Namely, they assess the impact of high bond market illiquidity, i.e., high VIX index, and fund liquidity, estimated as fund cash assets. In summary, while Chen and Qin emphasize a broad range of macro factors, Kuong et al. propose the outflow- Δ FFTar sensitivity mechanism with its direct connection to monetary policy, and Goldstein et al. explore how bond market volatility exacerbates fund flow dynamics during periods of economic stress. All provide valuable background for the cash ratio, FEDFUNDS and FEDFUNDS-CHG interpretations.⁵⁴

3.1.2. Influence of Fund Characteristics on Fund Flows

The relationship between fund flows and performance is another key area of investigation in these studies. Chen and Qin find that investor flows chase past fund performance but determine the flow-performance relationship to be non-convex and thus differ from that of equity funds. Also, they conclude that investor flows also predict fund performance. Goldstein et al. similarly study flow-performance relation and find it to be more concave, which implies investor outflows are more sensitive to inferior performance than inflows are to good fund performance. They confirm this finding's robustness across fund age, aggregate fund flow level, fund, and month fixed effects. Together, these studies show that investor behavior in CBMFs is highly sensitive to poor performance, which can decrease a fund's liquidity as investors leave.

The fund and market liquidity also play a pivotal role across fund flow studies. Chen and Qin incorporate aggregate market liquidity factors, i.e., VIX, and find that it positively correlates with CBMF flows. However, fund-level liquidity is not integrated into their analysis. Goldstein et al. examine the influence of fund and market liquidity on investor flows. They find that illiquidity exacerbates fund outflows on both the fund and market level. Also, they discuss the resulting potential financial fragility of markets and suggest remedies in the form of cash buffers or regulations. Kuong et al. explore liquidity in further regressions which indicate that market illiquidity and the staleness of fund share prices amplify the outflow- Δ FFTar sensitivity, which they consider to be novel findings. Thus, they determine a MPTC affecting corporate bond fund flow patterns. Together, these studies emphasize the significance of liquidity in CBMF dynamics. Goldstein et al. highlight the risks of liquidity mismatches, while Kuong et al. provide novel insights into how liquidity interacts with monetary policy to shape investor behavior.

3.2. Impact of COVID-19 on Fund Flows and Corporate Bond Markets

The COVID-19 crisis profoundly affected corporate bond fund flows, with studies examining how flows evolved during this period and identifying key factors driving these changes. Falato et al. examine corporate bond fund flows during and in the wake of the COVID-19 crisis, which they define as the period from February to April 2020. They use a regression analysis to estimate the relationship between fund flows and the various pandemic stages by utilizing the chronological dummy variables Crisis (February – April 2020), Peak (March 13 – 23, 2020), First Response (March 23 – April 9, 2020) and Second Response (April 9 – 17, 2020). The latter two dummies refer to the Fed policy announcements regarding the PMCCF and SMCCF on March 23 and their endowment with \$75 billion on April 9. Outflows peaked during mid-to-late March, the height of the crisis. The authors show that, on average, funds experienced an unprecedented 10% cumulative outflow. Then, they explore the effect of and the mechanism behind the Fed policy announcements and bond purchase program, which improved fragility by supplying a liquidity backstop in a liquidity drained market. They conclude with an analysis of the Fed's lasting effect on flows and liquidity after the crisis (April – August 2020), finding that cumulative inflows rebounded remarkably quickly to an average of 9%. Afterward, Falato et al. assess sources of fragility and identify three main sources: asset illiquidity, fire-sales vulnerability, and sector exposure. They find that illiquid funds experienced preemptive and more extreme outflows relative to other funds.⁵⁵

Boyarchenko et al. similarly examine fund flows during this period but focus on the effect of the Federal Reserve Bank of New York corporate credit facilities, i.e., PMCCF and SMCCF on corporate bond markets in the wake of the COVID-19 pandemic.⁵⁶ The PMCCF was created as a “funding backstop” and the SMCCF as a provider of “market liquidity for corporate bonds”.⁵⁷ Ultimately, they find that aggregate liquidity on the secondary debt market was improved mainly through announcement effects.⁵⁸

O'Hara and Zhou investigate the COVID-19 corporate bond liquidity crisis, which they define as the liquidity crunch and its repercussions for liquidity and transaction costs and bond pricing in the weeks before the PMCCF and SMCCF creation. They find that the interventionist actions established the Fed as “market maker of last resort, . . . willing to buy assets directly or to facilitate such buying by taking such assets as collateral” who injected liquidity back into the corporate bond market. This market improvement was already effective through announcements alone, which preceded the actual implementation, leading to a speedy resolution of the liquidity crisis.⁵⁹ All three papers agree that the Fed's an-

⁵⁵ cf. Falato et al. (2021), pp. 35–52

⁵⁶ cf. Boyarchenko et al. (2022), pp. 695–731

⁵⁷ cp. Federal Reserve Bank of New York (2020), URL see References

⁵⁸ cf. Boyarchenko et al. (2022), p. 707

⁵⁹ cf. O'Hara and Zhou (2021), pp. 46–68

⁵³ cf. Kuong et al. (2024), pp. 1–19

⁵⁴ cf. Goldstein et al. (2017), pp. 592–613

nouncements significantly calmed investors and improved corporate bond market stability. Together, these findings demonstrate the scale of outflows during the COVID-19 crisis and the pivotal role of the Fed's interventions in restoring stability to corporate bond markets.

3.3. Datasets and Methodologies in Fund Flow Studies

The datasets and methodologies employed by these studies vary and provide valuable suggestions for this thesis's analysis. Chen and Qin construct their dataset from US corporate bond funds information from 1991-2014 and cover 418 unique bond funds, consisting of 229 HY and 189 IG bonds. They then conduct fund-level time-pooled and cross-sectional regressions of the net fund flows on the fund characteristics controls and macro variables. The latter consist of the bond-related BOND, STK, DEF, OPTION, TB3 and VIX factors.⁶⁰

Goldstein et al. use a larger data sample consisting of corporate bond data from January 1992 to December 2014 with 4679 unique fund share classes and 1660 unique corporate bond funds. Their fund-level analyses regress the corporate bond's net flow on the fund alpha variable and fund characteristics controls using multiple estimations, enabling them to study flow-performance dynamics and liquidity mismatches in depth.⁶¹

Kuong et al. use a more recent dataset of 3182 unique funds and 6251 unique share classes, covering the years 2009 to 2023, composed of FfTar from the Federal Reserve Economic Data (FRED), Federal Open Market Committee dates and Futures related information. Their analyses include event studies and cross-sectional time-series regressions, allowing them to capture the impact of FfTar changes and crises like COVID-19 on fund flows.

Falato et al. construct their dataset from US corporate bond high-frequency real-time daily fund flows and returns data from January 2010 to April 2020 as well as fund characteristics information, all from Morningstar. In summary, they observe 1511 unique funds and 1511 unique share classes. First, they determine the scale of the impact on funds flows during the pandemic by documenting the change in flow patterns during the crisis in graphical and statistical analyses. Then, they conduct a multiple regression analysis with crisis-stage dummy variables to study the evolution of fund flows during the pandemic and the effectiveness of Fed policy actions.⁶²

Boyarchenko et al. use bond issuance, dealer, trade, and daily bond information spanning 2020. They regress various spreads on the secondary market against differently-rated corporate bonds and dummy variables indicating the timeline around facility announcements. Their analysis highlights the effects of the PMCCF and SMCCF on differently rated bonds,

emphasizing the role of announcement effects in stabilizing the market.⁶³

O'Hara and Zhou assemble their sample from corporate bond transaction data, characteristics, and dealer information, spanning February 1, 2020, to May 19, 2020. They analyze the microstructure of liquidity provision by measuring the corporate bond illiquidity proxied by trading transaction costs across bond categories. Moreover, they study primary and non-primary dealers and their ability and efficaciousness as liquidity providers. They employ a variety of regressions with trading volume, dealer inventory changes, transaction costs and market liquidity as dependent variables and control for industry and trade size fixed effects.⁶⁴

3.4. Formulation of Research Question

Chen and Qin focus on only a select few macro factors. Furthermore, Kuong et al. elaborate on this question by focusing on the impact changes in the Federal Funds Target rate while Goldstein et al. focus on the influence of general fund characteristics and analyze the flow-performance relationship. Building on this, the present thesis seeks to reduce that research gap by focusing on the effective federal funds rate and rate change. These factors present a novel opportunity to examine two macro factors which are closely tied to the Fed.

While Falato et al. study the evolution of corporate bond fund flows during the COVID-19 crisis, Boyarchenko et al. examine the impact of the Fed interventions on credit spreads in corporate bond markets. Additionally, O'Hara and Zhou analyze the liquidity provision during the COVID-19 crisis. This thesis builds on and expands their findings by focusing on the medium-term impact of COVID-19 on CBMF flows until December 2021. Moreover, the cash ratio could be another key variable. As a potential liquidity buffer, it may protect CBMFs against investor withdrawals and associated liquidation costs. Hence, I plan to analyze whether investors reward fund-level liquidity, inspecting the correlation between the cash ratio and fund flows.

Like Chen and Qin, this thesis employs monthly cross-sectional and time-pooled regressions, with and without time and fund style fixed effects. The same macro variables are included (BOND, VIX, STK, DEF, OPTION and TB3). Also, I use subsets of variables to improve the robustness of my findings. Once again similar to Chen and Qin, I differentiate between fund styles in my analysis, but conversely to them, I analyze every Morningstar Category to extract more nuanced fund style results. Following Falato et al. and O'Hara and Zhou, this thesis incorporates time dummy variables and interaction terms in the COVID-19 crisis case study to delineate the pandemic phases. This analysis includes interaction terms with the cash ratio, the performance rank, and changes in the effective federal funds rate. In summary, this thesis examines the determinants of CBMF flows with a special focus

⁶⁰ cf. Y. Chen and Qin (2017), pp. 2 ff.

⁶¹ cf. Goldstein et al. (2017), p. 598 ff.

⁶² cf. Falato et al. (2021), p. 38 ff.

⁶³ cf. Boyarchenko et al. (2022), p. 695 ff.

⁶⁴ cf. O'Hara and Zhou (2021), p. 51 ff.

on the effective federal funds rate, rate changes, and cash ratio. Then it proceeds with a fund style split sample analysis. Finally, it investigates the medium-term effects of the COVID-19 crisis on the CBMF flows in a dedicated case study.

4. Data and Methodology

This chapter outlines the data sources and methodology used for the analysis of CBMF flows. The primary dataset is monthly US CBMF information from Morningstar (2001-2021). Additionally, macro factors are extracted from Bloomberg and FRED databases. The methodology focuses on constructing the main variables. The chapter also discusses the descriptive statistics of both the fund and macro data.

4.1. Corporate Bond Mutual Fund and Macroeconomic Data

The primary data is monthly US corporate bond mutual fund information from Morningstar: fund identification, share class identification, total raw returns, inception date, Morningstar category, asset allocation, expense ratio, date, and net assets, with the data spanning the years from August 2001 to December 2021. Fund turnover information was included initially, but its poor data availability reduced the sample by tens of thousands of observations, which would have weakened the analysis. Multiple comparative studies also proceed without turnover data.⁶⁵ Consequently, the data was removed from this analysis. The original data includes 721 unique funds and 2814 unique share classes. In the final sample, share classes are aggregated to fund level, leaving 681 unique funds. As funds consist of share classes with unique flows, returns and expense ratios, I use asset-weighted net fund flows, total returns and net expense ratios by calculating weighted averages of the aggregated share classes relative to the corresponding fund's total net assets (TNA).⁶⁶ I then perform a monthly 90% winsorization of the flows, returns and expense ratios by winsorizing them at the 5th and 95th percentiles. The final cumulative net flow is plotted in Figure 3. Noticeably, the funds in the sample received inflows of tens of billions of US dollars which aligns with the expanding US bond market.

Each fund has a minimum average bond asset allocation of 40% over the sample period and is thus classified as a CBMF. The cash ratio, bond ratio, equity ratio and other ratio (the remaining percentage) are taken from the asset allocations. The cash ratio, bond ratio and equity ratio over time are visualized in Figure 4. The cash and bond lines show that CBMFs hold the majority of their assets in bonds and only a small percentage in cash. Of the bond holdings, roughly 40% consist of corporate bonds, however, the median corporate bond ratio sways between 50% to 70% percent (see

Figure 4). This negative skew suggests that some funds hold fewer corporate bonds, lowering the average.

The expense ratio is backfilled for the preceding months of the year since funds report it at end of the fiscal year. The fund age is calculated as the difference between the current date and the fund inception age in years, under the assumption that the minimum inception age applies to all share classes.⁶⁷ Funds differ in their reporting frequency: Some report their current asset allocation monthly, while others report each quarter. Therefore, the gaps between reports are filled by assigning the prior date's asset allocation forward until the next date, unless the next date is the last date. The volatility is the current date's standard deviation of the cumulative monthly returns in the prior 12 months. Finally, Morningstar assigns each Morningstar Category to each fund. The nine Morningstar Categories are (in decreasing sample frequency): US Fund High-Yield Bond (280 funds), US Fund Intermediate Core Bond (111 funds), US Fund Short-Term Bond (102 funds), US Fund Multisector Bond (70 funds), US Fund Corporate Bond (54 funds), US Fund Intermediate Core-Plus Bond (47 funds), US Fund Nontraditional Bond (29 funds), US Fund Long-Term Bond (29 funds), and US Fund Bank Loan (9 funds).

Table 1 shows the summary statistics for the sample's fund characteristics. The average fund has \$1.357 billion total net assets, is 14.424 years old, with an asset allocation of 88.459% bonds and 7.558% cash. Its net inflows are 0.469% with 0.421% total returns, 1.382 volatility and a net expense ratio of 0.823%. In comparison, the median fund has \$298.392 million total net assets, is 12.217 years old and allocates more assets to bonds with a 91.661% bond ratio and only 4.806% cash ratio.

It experiences a slight net outflow at the median but has higher returns and a lower volatility and net expense ratio. This suggests a positive skew of the total net assets, fund age, volatility, net expense ratio and cash ratio. The percentiles of the total net assets show a heterogeneous, wide distribution. In comparison, the fund age difference of 2.207 years is relatively small. The distribution of fund flows includes firmly positive and negative percentiles, which indicates quite heterogeneous fund flows. More than half of the funds must experience small-to-substantial net outflows. The total returns and bond ratio are negatively skewed. The contrast between the mean fund and median fund is unexpected as it implies that investor flows may not conform to traditional risk-return tradeoff expectations. However, the standard deviations of the flows, volatility, expense ratio and returns encompass these ratios. Consequently, the sample is considered robust enough for further analysis.

The macro data is drawn from both the Bloomberg and FRED databases. The Barclays aggregate bond index, intermediate government bond index, US corporate high yield index, S&P 500 index and GNMA index are extracted from Bloomberg. The indices here and hereafter refer to the to-

⁶⁵ cf. Choi et al. (2020), p. 435, cf. Y. Chen and Qin (2017), p. 2 f., cf. Goldstein et al. (2017), p. 598, cf. Q. Chen et al. (2010), p. 244 f.

⁶⁶ cf. Pástor et al. (2015), p. 31

⁶⁷ cf. Pástor et al. (2015), p. 31

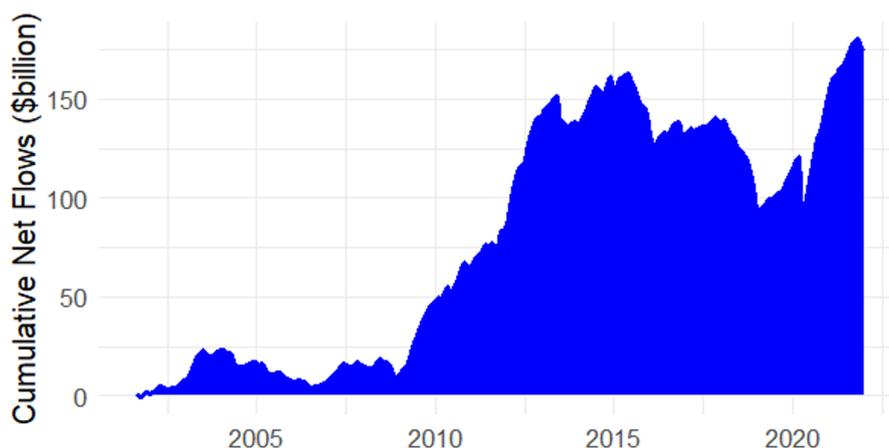


Figure 3: Cumulative net flow of the data sample in percent

This graph plots the cumulative net flows of all funds in the sample in the sample period, August 2001 to December 2021.

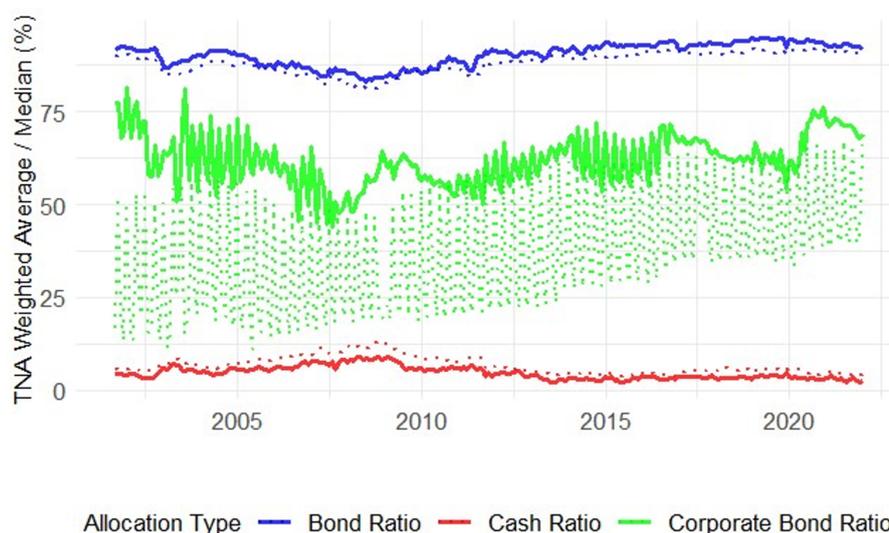


Figure 4: TNA weighted average and median of the sample cash, bond, and corporate bond ratios

This graph plots the time series of the sample's bond ratio, cash ratio and corporate bond ratio. The solid line shows the TNA-weighted median, and the dotted line represents the TNA-weighted average.

tal return index with gross dividends. The FEDFUNDS, effective federal funds rate change (FEDFUNDS-CHG, in percent), three-month Treasury bill secondary market rate (TB3MS) and four-week Treasury bill secondary market rate are downloaded from FRED. The market volatility index (VIX), "measuring implied volatilities in stock market", is obtained from the Chicago Board Options Exchange.⁶⁸ Since the volatility index is reported daily, each month's arithmetic VIX average is used.

4.2. Construction of Main Variables

The fund flow is the main dependent variable. The independent variables include both fund characteristics and

macro factors. All independent variables are lagged by one month. The fund controls include the logarithm of fund age, of fund size, i.e., total net assets, as well as the volatility, expense ratio, the past month's flows, and the performance rank. The performance rank is a fund's fractional rank based on its cumulative monthly returns in the prior 12 months relative to other funds in the same Morningstar Category.⁶⁹ The better a fund's percentile performance, the higher its assigned rank, ranging from 0 to 1. The cash asset allocation ratio held by CBMFs represents the approximate fund-level liquidity.⁷⁰

⁶⁸ cp. Goldstein et al. (2017), p. 595

⁶⁹ cf. Y. Chen and Qin (2017), p. 5

⁷⁰ cf. Goldstein et al. (2017), p. 599

Table 1: Descriptive statistics: fund characteristics

Variable	N	Mean	SD	P5	P25	P50	P75	P95
Total net assets [\$ million]	93198	1357.417	3758.315	12.721	78.593	298.392	1017.033	6116.023
Fund age [year]	93198	14.424	11.000	1.000	5.610	12.217	20.660	35.306
Net fund flows [%]	92771	0.469	3.534	-4.481	-1.277	-0.011	1.619	7.303
Total returns [%]	92771	0.421	1.806	-2.008	-0.169	0.438	1.170	2.811
Volatility [%]	92263	1.382	1.110	0.279	0.701	1.100	1.661	3.579
Net expense ratio [%]	93198	0.823	0.383	0.148	0.607	0.800	1.035	1.474
Cash Ratio [%]	93197	7.558	9.058	0.258	2.366	4.806	9.236	24.101
Bond Ratio [%]	93197	88.459	11.214	67.156	85.564	91.632	95.391	98.685
Equity Ratio [%]	93197	0.889	3.528	0.000	0.000	0.000	0.325	4.390
Other Ratio [%]	93197	3.092	5.060	0.000	0.141	1.420	3.968	11.059

This table shows the summary statistics for the sample's fund characteristics: N, mean, standard deviation, and the 5th, 25th, 50th (median), and the 95th percentiles.

Additionally, the BOND, DEF, OPTION and STK factors are calculated. BOND is the return spread between the Barclays aggregate bond index and the 4-week Treasury bill rate, DEF is the "return spread between the high-yield bond index and the intermediate government bond index", OPTION is the "return spread between the GNMA index and the intermediate government bond index" and STK is the stock market return spread between the S&P 500 index and the 4-week Treasury bill rate.⁷¹ Finally, I use a second liquidity measure: The aggregate corporate bond market illiquidity is proxied by the implied market volatility index, VIX.⁷² I assign the BOND, DEF, OPTION, STK and VIX factors as macro controls. Table 2 shows the summary statistics for the macro factors. On average, the BOND factor is -0.779%, the FEDFUNDS is 1.269 with a rate change of -0.012%, the VIX index is 19.254, the STK factor is -0.241%, the DEF factor is 0.357%, the OPTION factor is 0.046% and the three-month Treasury bond index is 1.162. At the median, the BOND factor is -0.483%, the FEDFUNDS is 0.650 with a rate change of 0.000%, the VIX index is 17.273, the STK factor is -0.042%, the DEF factor is 0.566%, the OPTION factor is 0.093% and the three-month Treasury bond index is 0.510.

5. Analysis and Results

This chapter outlines the analysis of the fund flow determinants using regression models and the results, focusing on both systematic dynamics and the impact of the COVID-19 pandemic. First the systematic fund flow regressions are presented. Next, a comprehensive fund style split sample analysis is conducted, and the results are compared. Finally, the impact of the COVID-19 on CBMF flows is investigated.

⁷¹ cp. Y. Chen and Qin (2017), p. 5

⁷² cf. Goldstein et al. (2017), p. 599, cf. Kuong et al. (2024), p. 15

5.1. Systematic Fund Flow Dynamics: Regression Models and Results

I estimate the influence of the fund characteristics and macro factors with pooled time-series and cross-sectional linear regressions, namely OLS regressions without fixed effects and FE regressions with fixed time and fund style effects similar to Chen and Qin.⁷³ The fund flow remains the dependent variable for all regressions. The time fixed effects control for the year and month, and the fund style fixed effect controls for the Morningstar Category. The effective federal funds rate, its change, the VIX index, the BOND factor, the cash lag, and the performance rank are of particular interest. The remaining fund characteristics, i.e., the logarithm of the fund age and total net assets, the return volatility, the expense ratio and the previous month's flows are fund control variables. Similarly, the three-month Treasury bill rate and the STK, DEF, OPTION factors are macro controls. These variables were selected to estimate the fund flow correlations accurately as they are standard in published fund flow studies like Elton et al. and Chen and Qin.⁷⁴ Using heteroskedastic and autocorrelation consistent standard errors is also the standard in published flow studies. This is because financial market time series data is known to exhibit heteroskedasticity and have time-varying standard deviations for predictor variables.⁷⁵ Additionally, financial market time series data often exhibits autocorrelation. For example, fund returns have been shown to be positively correlated.⁷⁶ Furthermore, daily fund flows have also been proven to be significantly autocorrelated.⁷⁷ Consequently, a studentized Breusch-Pagan-

⁷³ cf. Y. Chen and Qin (2017), p. 6f.

⁷⁴ cf. Elton et al. (1995), 1234, 1241, cf. Y. Chen and Qin (2017), p. 8

⁷⁵ cf. Investopedia (2025), URL see References

⁷⁶ cf. Choi et al. (2020), p. 297

⁷⁷ cf. Edelen and Warner (2001), p. 199 f., cf. Rakowski and Wang (2009), 2104 f.

Table 2: Descriptive statistics: macroeconomic condition

Variable	N	Mean	SD	P5	P25	P50	P75	P95
BOND [%]	87160	-0.779	1.606	-4.191	-1.593	-0.483	0.317	1.309
FEDFUNDS [%]	87160	1.269	1.501	0.080	0.120	0.650	1.910	5.020
FEDFUNDS-CHG [%]	87160	-0.012	0.158	-0.280	-0.010	0.000	0.020	0.170
VIX	87160	19.254	8.242	11.062	13.678	17.273	22.374	34.05
STK [%]	87160	-0.241	4.554	-8.795	-2.680	-0.042	2.422	6.846
DEF [%]	87160	0.357	2.874	-3.522	-0.556	0.566	1.485	4.229
OPTION [%]	87160	0.046	0.428	-0.743	-0.182	0.093	0.309	0.694
TB3MS [%]	87160	1.162	1.400	0.020	0.080	0.510	1.760	4.720

This table shows the summary statistics for the macroeconomic factors during the sample period: N, mean, standard deviation, and the 5th, 25th, 50th (median), and the 95th percentiles.

Test is performed for each regression. The test has two hypotheses: The null hypothesis is that homoscedasticity is present, the complementary hypothesis is that heteroskedasticity is present. If the test statistic output's ("BP") corresponding p-value indicates significance, the null hypothesis must be rejected, and heteroskedasticity is present. Therefore, OLS standard errors may be less dependable. Consequently, Newey-West standard errors are calculated, which have the advantage of being both heteroskedasticity and autocorrelation consistent.

First, I analyze the influence of the macro condition on the fund flows by performing iterative regressions. In OLS regression 1, I regress the fund flows on the fund and macro variables (see equation 2). α represents the regression intercept, β the coefficients and $\varepsilon_{i,t}$ the residual error term.

$$\begin{aligned} Flows_{i,t} = & \alpha + \beta_1 \times cash\ ratio_{i,t-1} \\ & + \beta_2 \times FEDFUNDS-CHG_{i,t-1} \\ & + \beta_3 \times performance\ rank_{i,t-1} \\ & + \beta_4 \times Controls_{i,t-1} + \varepsilon_{i,t} \end{aligned} \quad (2)$$

In OLS regression 2, I repeat the regression while limiting the independent variables to the macroeconomic factors (see equation 3).

$$\begin{aligned} Flows_{i,t} = & \alpha + \beta_1 \times FEDFUNDS-CHG_{i,t-1} \\ & + \beta_2 \times Controls_{i,t-1} + \varepsilon_{i,t} \end{aligned} \quad (3)$$

In FE regression 3, I conduct a regression on all independent variables while controlling for style fixed effects (see equation 4). η_i describes the control term for style fixed effects.

$$\begin{aligned} Flows_{i,t} = & \alpha + \beta_1 \times cash\ ratio_{i,t-1} \\ & + \beta_2 \times FEDFUNDS-CHG_{i,t-1} \\ & + \beta_3 \times performance\ rank_{i,t-1} \\ & + \beta_4 \times Controls_{i,t-1} + \eta_i + \varepsilon_{i,t} \end{aligned} \quad (4)$$

Time fixed effects are excluded as the macro conditions primarily change with transitory time and hence could lose informative value. Performing the Breusch-Pagan test reveals highly significant results for all three regressions (see Table 3) with p-values $< 2.2e^{-16}$. Accordingly, the null hypothesis is rejected, heteroskedasticity is present, and I calculate Newey-West standard errors. Then, I further analyze the influence of the fund characteristics on the fund flows by performing iterative regressions. In the FE Regression 4, I regress the fund flows on the fund variables while controlling for style and time fixed effects (see equation 5). λ_t is the control term for the time fixed effects.

$$\begin{aligned} Flows_{i,t} = & \alpha + \beta_1 \times cash\ ratio_{i,t-1} \\ & + \beta_2 \times performance\ rank_{i,t-1} \\ & + \beta_3 \times Controls_{i,t-1} + \eta_i + \lambda_t + \varepsilon_{i,t} \end{aligned} \quad (5)$$

Table 3: Breusch-Pagan test results for the regression models 1-4

Model	Statistic	p-value
Regression 1	4092.193***	$< 2.2e^{-16}$
Regression 2	2518.826***	$< 2.2e^{-16}$
Regression 3	4666.439***	$< 2.2e^{-16}$
Regression 4	8261.537***	$< 2.2e^{-16}$

This table shows the Breusch-Pagan test results for regressions 1-4. Stars signify statistical significance: *** p < 0.001, ** p < 0.01, * < 0.05, respectively.

I also conduct Breusch-Pagan tests (see Table 3) with highly significant results (p-value $< 2.2e^{-16}$) that confirm heteroskedasticity and calculate the Newey-West standard errors. I then proceed with the different regression analyses. Table 4 presents the results.

The expected correlation between FEDFUNDS and FEDFUNDS-CHG to fund flows is negative as a high FEDFUNDS

Table 4: OLS and style FE regressions 1-4

	Dependent variable: flows			
	All vars.	Macro. vars.	All vars.	Fund vars.
	OLS	OLS	Style FE	Time FE
	<i>Newey-West</i>	<i>Newey-West</i>	<i>Newey-West</i>	<i>Newey-West</i>
	(1)	(2)	(3)	(4)
volatility	-0.056*** (0.016) t = -3.379		-0.009 (0.019) t = -0.480	-0.113*** (0.022) t = -5.060
log(fund age)	-0.414*** (0.019) t = -21.295		-0.427*** (0.020) t = -21.400	-0.408*** (0.019) t = -21.292
past flows	0.372*** (0.008) t = 47.526		0.369*** (0.008) t = 47.123	0.373*** (0.008) t = 48.629
log(TNA)	0.017* (0.008) t = 2.061		0.015 (0.009) t = 1.726	0.016 (0.008) t = 1.956
expense ratio	0.003 (0.048) t = 0.069		0.090 (0.052) t = 1.749	0.006 (0.049) t = 0.118
performance rank	1.077*** (0.046) t = 23.229		1.086*** (0.047) t = 23.317	1.079*** (0.045) t = 23.757
cash ratio	0.005** (0.002) t = 2.642		0.003 (0.002) t = 1.667	0.005** (0.002) t = 2.779
BOND	0.153*** (0.012) t = 12.441	0.212*** (0.013) t = 15.853	0.153*** (0.012) t = 12.451	
FEDFUNDS	-0.266*** (0.070) t = -3.777	-0.204 (0.106) t = -1.921	-0.250*** (0.070) t = -3.560	
FEDFUNDS-CHG	0.046 (0.110) t = 0.416	0.030 (0.143) t = 0.208	0.035 (0.110) t = 0.316	

This table reports the fund-level regression results for the coefficient estimates of the independent variables. Fund flow is the dependent variable while fund characteristics and macro factors are the independent variables. Column (1) shows the OLS regression results for all variables. Column (2) shows the OLS regression estimates for macro variables. Column (3) shows the FE fund style regression for all variables. Column (4) shows the FE time regression for the fund variables. "Time FE" denotes time, i.e., year and month, fixed effects, and "Style FE" above Column (3) denotes fund style fixed effects. The Newey-West standard error sits below the coefficient in parentheses. The t-value, designated "t", is below it. Stars indicate statistical significance: *** p < 0.001, ** p < 0.01, * < 0.05, respectively.

Table 4 — continued

STK	0.019*** (0.005) t = 4.217	0.014** (0.005) t = 3.065	0.019*** (0.004) t = 4.191	
VIX	0.051*** (0.003) t = 18.243	0.056*** (0.004) t = 14.206	0.048*** (0.003) t = 16.622	
TB3MS	0.412*** (0.078) t = 5.304	0.405*** (0.117) t = 3.469	0.402*** (0.078) t = 5.178	
DEF	0.009 (0.008) t = 1.185	0.100*** (0.009) t = 11.133	0.006 (0.008) t = 0.769	
OPTION	0.089** (0.031) t = 2.882	0.067 (0.035) t = 1.885	0.085** (0.031) t = 2.759	
Intercept	-0.648*** (0.168) t = -3.857	-0.831*** (0.080) t = -10.360	-0.886** (0.339) t = -2.616	1.515*** (0.280) t = 5.408
Observations	87,160	87,160	87,160	87,160
R ²	0.206	0.027	0.208	0.252
Adjusted R ²	0.206	0.027	0.208	0.250

and its increases exacerbate liquidity costs by making borrowing money more costly.⁷⁸ Thereby, investor flows into funds should decrease.⁷⁹ As hypothesized, FEDFUNDS is consistently negative but varies in significance: highly significant in regression 1 and 3, but not significant in regression 2.

This implies that investors invest less during high effective federal rate conditions, which is highly economically plausible and significant as the factor loadings amount to roughly half of the average net fund flows of 0.469%. The decreased net flows across fund styles could suggest divestment from CBMFs of all fund styles when the FEDFUNDS rises, potentially indicating fire sales.⁸⁰ In contrast, the FEDFUNDS-CHG and flow relationship is positive, small and statistically insignificant across regressions and style fixed effects. Comparing this to Kuong et al., where the ΔFFTar correlates strongly and significantly (1.276 at the 1% level) with monthly bond outflows, offers two key insights:⁸¹ First, they support the hypothesis that investors closely consider macro factors tied to the FEDFUNDS. Second, they underscore the importance of the Fed as monetary policy maker

⁷⁸ cf. Goldstein et al. (2017), p. 611 f.

⁷⁹ cf. Kuong et al. (2024), p. 5

⁸⁰ cf. Falato et al. (2021), p. 3145

⁸¹ cf. Kuong et al. (2024), p. 6

and its transmission channels between corporate bond markets and monetary policy.⁸² Unlike Kuong et al., who determine the ΔFFTar to be the key factor, this analysis finds that FEDFUNDS, and not FEDFUNDS-CHG, is more significant.⁸³

The DEF coefficients are positive, slightly above zero, and not significant, except in regression 2, where it is 0.100 and highly significant. These results agree with Chen and Qin, who also find the DEF factor to be consistently small and positive for all funds and IG funds.⁸⁴ A positive DEF is economically plausible, as more investors tend to engage in flight-to-safety behavior during periods of higher default risk, favoring bonds over stocks.⁸⁵ The implied market volatility index coefficients 1-3 are highly significant with only minimal variation: 0.051, 0.056 and 0.048. Chen et al. receive similar results and Kuong et al. estimate $\Delta\log(\text{VIX}) = -0.179$ which equals $\text{VIX} \approx 0.836$.⁸⁶ Akin to investor behavior during times of higher default risk, the risk associated with higher stock market volatility, i.e., higher VIX, might also cause investors to prefer bond investments and lead to inflows.⁸⁷

⁸² cf. Kuong et al. (2024), p. 5

⁸³ cf. Kuong et al. (2024), p. 6

⁸⁴ cf. Y. Chen and Qin (2017), p. 8

⁸⁵ cf. Y. Chen and Qin (2017), p. 7 f.

⁸⁶ cf. Y. Chen and Qin (2017), p. 8, cf. Kuong et al. (2024), p. 6

⁸⁷ cf. Y. Chen and Qin (2017), p. 7 f.

The BOND coefficients are highly significant, ranging between 0.125 (regression 1 and 3) and 0.212. This is reasonable as higher excess returns on the aggregate bond index result in higher investor margins and are therefore more profitable, thereby attracting investors and net inflows across all fund styles. This finding is validated by Chen and Qin, who calculate BOND factors close to 0.2 for all, IG and HY funds.⁸⁸ The other macro controls, STK (significant to highly significant), OPTION (not significant to significant) and TB3MS (highly significant) have a consistently positive correlation to flow. These macro correlations are corroborated by Chen and Qin; however, the positive TB3MS correlation diverges from their findings as it would imply that rising interest rates, which lower the comparative bond yield spread, still lead to fund inflows.⁸⁹

Funds with higher cash ratios, i.e., better liquidity, should attract inflows from investors who value asset liquidity due to presumably lower liquidation and redemption costs. In fact, Chen et al. show that funds with lower liquidity tend to conduct costly liquidations to satisfy withdrawals.⁹⁰ Depending on the priorities of the overall investor population, I expect the correlation between cash ratio and fund flows to be positive to neutral. This expectation is met as the regression coefficients are near zero. This might indicate that flows are largely unrelated to the cash ratio, however, considering their varied significance, a firm conclusion cannot be drawn. While this suggests a limited economic significance of the cash ratio due to the coefficients' exceedingly small size, the correlation is statistically significant in some regressions (e.g., 0.005 in regression 1 and 0.003 in regression 3, though not significant). This aligns with Jiang et al., who suggest that funds often scale down liquid and illiquid holdings proportionally during financial strain to maintain their portfolio ratio, making cash ratio changes largely irrelevant to investors.⁹¹

The association of the fund controls log(fund age), log(TNA), past flows and performance rank is expected to be negative, negative, positive and positive, respectively.⁹² The past flow and especially the performance rank should be highly significant and have a larger effect as investors are hypothesized to strongly prioritize fund performance and recent flows. These correlations are confirmed by the results across regressions, style, and time fixed effects. The log(fund age), past flows and performance rank correlations behave very consistently and are economically significant due to their large size relative to the fund flows average and median.

The performance rank coefficients are highly significant and noticeably large in regressions 1, 3 and 4, i.e., the regressions involving all or the fund variables: They exceed 1.

As such, they are larger than the results the studies of Chen and Qin and Chen et al. (0.77 and 0.016), however, they remain below the Rakowski and Wang performance rank coefficient, 1.693.⁹³ Overall, this implies strong performance chasing from fund investors as the fund flows increase by more than 1 for every one-unit increase in the performance rank. This behavior is economically significant as the inflow is more than double the average fund flow, raising the question of investor overreactions.

The log(TNA) coefficient is marginally significant at 0.017 regression 1 but not robust to controlling for style or time fixed effects – in any case it is too small to be economically significant. This agrees with Goldstein et al., Rakowski and Wang and Chen and Qin.⁹⁴ The expense ratio correlations are slightly positive and not significant. Due to their lack of significance and large Newey-West standard errors, the results are not dependable. Otherwise, the positive correlation would be unusual as expenses were incorporated to control for the reportedly negative effect on fund flows.⁹⁵

The volatility is negative and generally highly significant as expected, indicating that increasing return volatility deters fund inflows which is validated by literature.⁹⁶ However, this result is not robust to controlling for style and too small to be economically significant. The adjusted coefficient of determination, adjusted R², shows how much of all variation in fund flows is explained by the incorporated variables. When employing all variables, it is broadly 21% (regression 1). It falls to 2.7% when the regression is confined to macro variables and controlled for style (regression 2). The maximum explanatory power is 25.2% when the regression is focused on fund variables and includes all FEs (regression 4). Broadly, this aligns with Chen and Qin who have adjusted R² values around 20%.⁹⁷

5.2. Systematic Fund Style Split Sample Analysis: Regression Models and Results

In FE regression 5, I conduct regressions on the fund variables and controls for each fund style by constructing individual and separate fund samples per Morningstar category while controlling for month fixed effects. The non-HY funds were not aggregated to an investment-grade sample because the goal is to examine the fund category's individual profile. The estimated model specification is (see equation 6):

$$\begin{aligned} Flows_{i,t} = & \alpha + \beta_1 \times cash\ ratio_{i,t-1} \\ & + \beta_2 \times performance\ rank_{i,t-1} \\ & + \beta_3 \times Controls_{i,t-1} + \eta_i + \lambda_t + \varepsilon_{i,t} \end{aligned} \quad (6)$$

⁸⁸ cf. Y. Chen and Qin (2017), p. 8

⁸⁹ cf. Y. Chen and Qin (2017), p. 7 f.

⁹⁰ cf. Q. Chen et al. (2010), p. 258

⁹¹ cf. Jiang et al. (2021), p. 1625

⁹² cf. Bergstresser and Poterba (2002), p. 405, cf. Y. Chen and Qin (2017), p. 8, cf. Sirri and Tufano (1998), p. 1599, cf. Pástor et al. (2015), p. 25

⁹³ cf. Y. Chen and Qin (2017), p. 6, 8, cf. Q. Chen et al. (2010), p. 248, cf. Rakowski and Wang (2009), p. 2107

⁹⁴ cf. Goldstein et al. (2017), p. 602, cf. Rakowski and Wang (2009), p. 2107, cf. Y. Chen and Qin (2017), p. 6

⁹⁵ cf. Sirri and Tufano (1998), p. 1612

⁹⁶ cf. Rakowski and Wang (2009), p. 227 f.

⁹⁷ cf. Y. Chen and Qin (2017), p. 6, 8

I confirm heteroskedasticity with Breusch-Pagan test (see Table 5) and calculate Newey-West standard errors. Table 6 presents the results of the regression results for each Morningstar Category. The factor loadings for the **cash ratio** range from slightly above zero to slightly below zero and are not significant across most Morningstar Category fund styles, i.e., short-term, intermediate core, intermediate core-plus, corporate, multisector, long-term and nontraditional bond funds.

This indicates that the cash ratio has no significant impact on fund flows in these categories, suggesting that these investors do not prioritize or reward larger cash buffers. There are multiple potential explanations for this finding. Short-term, intermediate core, intermediate core-plus, corporate, and long-term bond funds already primarily invest in investment-grade securities with low credit risk ratings. Hence, their investors might be less concerned with potential defaults, less likely to engage in flight-to-safety and thus less likely to suddenly withdraw. Nontraditional bond fund investors aim for specific objectives, e.g., absolute returns or lower volatility, which is why they likely deprioritize stability-focused measures such as the cash ratio. Another potential reason might be Jiang et al. showing that during times of financial stress, funds sell liquid and illiquid asset classes proportionally to maintain their original portfolio proportions.⁹⁸ Thus a higher cash ratio would not translate into lower internal redemption costs and be of little benefit for investors.

On the opposite end of the spectrum, the coefficients in high-yield bond funds and bank loan funds are marginally significant with 0.010 and 0.084, respectively. This implies a weak but positive association with the net fund flows with low economic significance due to their small size relative to the fund average of 0.469%. The explanation might be similar for both HY and bank loan funds. High-yield bond funds have an inherently bigger default and liquidity risk, as the underlying issues are rated as higher-risk securities with a bigger default risk and therefore are less easy to sell. During crises, they would be more vulnerable to fire sales and sudden redemptions.⁹⁹ Like high-yield bond funds, bank loan fund investments are also generally illiquid, as floating-rate debt is harder to shift due to its higher interest rate risk. This might motivate investors to allocate their funds in bank loan funds with larger liquidity buffers to safeguard their withdrawal rights and the funds' financial stability during economic downturns and lower interest rate environments.¹⁰⁰

Thus, investors might value HY bond and bank loan funds with larger cash ratios, since these liquidity buffers could mitigate their financial fragility in case of liquidity demand shocks by way of large redemptions or macro financial crises and economic downturns. However, Choi and Kronlund show that funds which tend to reach for yield, e.g., high-yield and bank loan funds, tend to practice less careful

liquidity management, i.e., hold fewer liquid assets.¹⁰¹ For this reason, the very small, marginally statistically significant and economically insignificant positive correlation could suggest that these fund investors do not place much importance on the cash ratio. Funds scaling down their liquid and illiquid assets proportionally offer a possible explanation.

Their magnitude is in line with Sirri and Tufano, who determine the performance rank coefficient to be 1.693, nearly twice as large as the rank coefficient results from Chen et al.¹⁰² The single outlier is the long-term bond fund category with the only negative and insignificant performance rank. Altogether, this clearly shows that performance rank and return momentum are a primary driver of fund flows: First, despite the typical investor heterogeneity of the fund styles, the majority of investors engage in significant performance chasing and allocate their capital in the recently high-performing funds. Second, the fact that the performance coefficient has the biggest effect of all variables shows that investors are highly sensitive to performance rank and prioritize it above all else. This conclusion is supported in prior studies from Chen et al. and Goldstein et al.¹⁰³ The largest coefficient is observed in nontraditional bond funds, which is plausible as nontraditional fund investors seek to fulfill specific goals in exchange for the atypical investment profile and are therefore immediately willing to shift if the fund disappoints their expectations. High-yield, short-term and multisector bonds all exhibit a similarly high investor responsiveness to performance rank which is logical as high-yield fund investors prioritize yield in exchange for higher credit and liquidity.

Additionally, this aligns with Goldstein et al. finding that illiquid funds have a higher performance sensitivity. This potentially applies to nontraditional, high-yield and multisector bonds, which are all considered higher risk.¹⁰⁴ Short-term and multisector funds attract investors more interested in liquidity and diversification, respectively, and potentially use high performance as a marker of quality or success. Interestingly, the intermediate core and core-plus performance rank coefficients differ significantly: While the former's is the lowest of all categories, the latter's is bigger by nearly half. This clearly shows that investors who chose intermediate-core plus funds with their bigger flexibility and higher credit and economic risk expect to be compensated with better performances and higher returns. Their sister category likely attracts more conservative investors who prioritize stability and other benchmarks compared to the other fund styles. The same can be hypothesized about corporate bond fund investors who have a similarly low performance coefficient at 0.634. Investors might value stability and believe in the long-term success of the underlying corporations or industries instead of being focused on immediate high returns.

The most surprising result is the small, negative, and not significant performance rank coefficient for long-term bond

⁹⁸ cf. Jiang et al. (2021), p. 1625

⁹⁹ cf. Choi and Kronlund (2018), p. 1959

¹⁰⁰ cf. Jiang et al. (2021), p. 1625

¹⁰¹ cf. Choi and Kronlund (2018), p. 1960

¹⁰² cf. Sirri and Tufano (1998), p. 1559, cf. Q. Chen et al. (2010), p. 248

¹⁰³ cf. Q. Chen et al. (2010), p. 247, cf. Goldstein et al. (2017), p. 605

¹⁰⁴ cf. Goldstein et al. (2017), p. 605

Table 5: Breusch-Pagan test results for the regression model 5 for each fund style

Split Sample Morningstar Category	Statistic	p-value
US Fund Corporate Bond	929.868***	$< 2.2e^{-16}$
US Fund Nontraditional Bond	554.967***	$< 2.2e^{-16}$
US Fund Multisector Bond	955.450***	$< 2.2e^{-16}$
US Fund High Yield Bond	4190.997***	$< 2.2e^{-16}$
US Fund Short-Term Bond	1517.569***	$< 2.2e^{-16}$
US Fund Intermediate Core Bond	866.684***	$< 2.2e^{-16}$
US Fund Long-Term Bond	555.690***	$< 2.2e^{-16}$
US Fund Intermediate Core-Plus Bond	853.275***	$< 2.2e^{-16}$
US Fund Bank Loan	323.488**	0.001

This table shows the Breusch-Pagan test results for regression 5 for each fund style as indicated by the Morningstar Category. Stars denote statistical significance: *** $p < 0.001$, ** $p < 0.01$ and * $p < 0.05$.

funds, -0.019. Long-term bonds have maturation durations of more than six years, making them more susceptible to interest rate risk and similar macro factors than comparable categories. Thus, their investors likely value long-term income stability instead of short-term performance and make their asset allocation decisions independent of the fund's relative rank.

The **fund age** correlations behave as expected for most bond fund categories, i.e., they are highly statistically significant and negative, ranging from -0.3 to -0.5. This implies that older high-yield, short-term, intermediate core, intermediate core-plus, corporate bond, and multisector bonds receive less net fund flows in a significant economic magnitude. Bank loan and nontraditional bond funds are the exception as both have statistically and economically less significant age coefficients, showing that fund age has no material impact on fund flows. Bank loan and nontraditional investors likely focus on other metrics like performance rank instead of age. They may also be more sophisticated in their niches and less likely to fall prey to recency and marketing biases. In sum, the $\log(\text{fund age})$ correlation is in line with Bergstresser and Poterba, Chen et al., Goldstein et al., Chen and Qin, and Rakowski and Wang, who also determine the association to be negative.¹⁰⁵

The **expense ratio** coefficients are near zero and not significant for high-yield, short-term, intermediate core, intermediate core-plus, corporate, bank loan and nontraditional bond funds, suggesting that expenses do not influence investor fund flow patterns, statistically, and would be too small to be economically significant even if they did. These results agree with Sirri and Tufano and the investment-grade and complete sample coefficients from Chen and Qin but are far from the findings of Goldstein et al. with -0.200.¹⁰⁶ Ad-

ditionally, Chen and Qin find a HY expense ratio correlation of -0.153, which is much bigger than this analysis's result, -0.002.¹⁰⁷

The outliers are long-term bond funds whose coefficient is -1.045 and multisector bond funds whose coefficient is 1.002. Both are highly statistically significant at the 0.1% level and very large, making them highly economically significant. Investors in long-term funds are likely highly averse to any factors that might erode their payments in addition to their high interest rate risk. In turn, it is plausible that they might also be highly cost-sensitive and penalize higher expense ratios harshly. Still, the extremely positive association between the expense ratio and net fund flows for multisector funds is atypical and economically implausible, as this would make expenses the second-strongest determinant of fund inflows after performance rank. Possibly higher expenses indicate a higher standard of quality regarding active managers and asset selections. However, such an excellent multisector bond manager could be expected to manage larger funds, resulting in a similarly significant and large fund size coefficient. This is not the case, as the $\log(\text{TNA})$ coefficient is insignificant, small, and negative. Furthermore, at an absolute value of 1, the long-term and multisector expense correlations exceed all comparative analyses. At this point, it would be prudent to repeat the analysis with a larger sample and larger sample period to assess the robustness of the result.

The effect of the **recent past flows** on the net fund flows is highly significant, positive, ranges between 0.2 and 0.5 and aligns with literature. This indicates a momentum or persistence in flows, i.e., past flows predict current flows. The more flexibility the fund style allows and the higher the focus on performance, the bigger the past flow coefficient is: Indeed, the top categories are bank loan, nontraditional, intermediate core-plus and multisector bond funds. Overall, the past flow coefficient magnitude is moderately larger compared to

¹⁰⁵cf. Bergstresser and Poterba (2002), cf. Q. Chen et al. (2010), p. 248, cf. Goldstein et al. (2017), p. 602, cf. Y. Chen and Qin (2017), p. 6, cf. Rakowski and Wang (2009), p. 2107

¹⁰⁶cf. Y. Chen and Qin (2017), p. 6

¹⁰⁷cf. Y. Chen and Qin (2017), p. 6

Table 6: Fund-specific time FE regression 5 for fund styles

Fund style	Dependent variable: fund flows								
	High-Yield Time FE Newey-West (1)	Short-Term Time FE Newey-West (2)	Intermediate Core Time FE Newey-West (3)	Intermediate Core-Plus Time FE Newey-West (4)	Corporate Bond Time FE Newey-West (5)	Multiset or Time FE Newey-West (6)	Long-Term Time FE Newey-West (7)	Nontraditional Time FE Newey-West (8)	Bank Loan Time FE Newey-West (9)
volatility	-0.519*** (0.071)	-0.376*** (0.113)	-0.211 (0.116)	-0.047 (0.128)	0.321* (0.127)	-0.034 (0.084)	-0.585* (0.269)	0.040 (0.152)	-0.361 (0.457)
log(fund age)	t = -7.273	t = -3.340	t = -1.824	t = -0.364	t = 2.532	t = -0.400	t = -2.179	t = 0.264	t = -0.791
	-0.419*** (0.033)	-0.370*** (0.047)	-0.413*** (0.054)	-0.385*** (0.074)	-0.520*** (0.065)	-0.567*** (0.052)	-0.329** (0.107)	-0.146 (0.130)	0.037 (0.304)
	t = -12.803	t = -7.806	t = -7.681	t = -5.177	t = -8.046	t = -10.894	t = -3.091	t = -1.124	t = 0.121
past flows	0.297*** (0.012)	0.377*** (0.016)	0.342*** (0.022)	0.496*** (0.026)	0.369*** (0.024)	0.478*** (0.023)	0.287*** (0.030)	0.496*** (0.035)	0.539*** (0.064)
	t = 25.455	t = 22.944	t = 15.567	t = 18.866	t = 15.569	t = 20.967	t = 9.730	t = 14.315	t = 8.445
log(TNA)	-0.008 (0.015)	0.035 (0.018)	0.016 (0.027)	0.030 (0.029)	0.090*** (0.026)	-0.011 (0.022)	-0.043 (0.039)	-0.094 (0.072)	-0.431* (0.185)
	t = -0.527	t = 1.892	t = 0.587	t = 1.027	t = 3.440	t = -0.489	t = -1.096	t = -1.319	t = -2.334
expense ratio	-0.002 (0.089)	0.059 (0.149)	-0.025 (0.137)	0.212 (0.189)	-0.013 (0.145)	1.001*** (0.179)	-1.045*** (0.299)	-0.244 (0.287)	0.666 (0.439)
	t = -0.022	t = 0.395	t = -0.184	t = 1.118	t = -0.089	t = 5.580	t = -3.496	t = -0.851	t = 1.516
performance rank	1.247*** (0.077)	1.248*** (0.121)	0.604*** (0.109)	1.107*** (0.123)	0.634*** (0.133)	1.248*** (0.125)	-0.019 (0.197)	1.327*** (0.235)	0.909*** (0.350)
	t = 16.279	t = 10.328	t = 5.538	t = 9.020	t = 4.764	t = 10.020	t = -0.094	t = 5.638	t = 2.598
cash ratio	0.010* (0.004)	0.002 (0.004)	0.006 (0.006)	-0.001 (0.006)	0.013 (0.008)	-0.005 (0.003)	0.025* (0.011)	0.002 (0.007)	0.084* (0.038)
	t = 2.381	t = 0.448	t = 1.150	t = -0.195	t = 1.788	t = -1.889	t = 2.353	t = 0.257	t = 2.219
Intercept	2.775*** (0.506)	2.127** (0.783)	1.111 (0.603)	2.927* (1.178)	0.060 (0.811)	2.206 (1.302)	3.569*** (0.945)	2.300 (1.506)	6.562 (3.948)
	t = 5.482	t = 2.717	t = 1.842	t = 2.485	t = 0.073	t = 1.694	t = 3.777	t = 1.528	t = 1.662
Observations	35,304	13,737	11,165	7,195	6,894	7,706	2,336	2,093	730
R ²	0.294	0.281	0.208	0.387	0.299	0.425	0.264	0.448	0.585
Adjusted R ²	0.289	0.268	0.189	0.365	0.273	0.406	0.175	0.373	0.369

This table reports the fund-level regression 5 results for the coefficient estimates of the fund specific variables. Fund flow is the dependent variable while fund characteristics and macro factors are the independent variables. "Time FE" above the columns denotes time, i.e., year and month, fixed effects. The Newey-West standard errors sits below the coefficient in parentheses. The t-value, designated "t", is below the standard deviation. Stars show statistical significance: *** p < 0.001, ** p < 0.01, * p < 0.05, respectively.

Goldstein et al. (0.152) and Chen et al. (0.14 – 0.24) and Chen and Qin who determine the lagged flow coefficients to be 0.210 for HY and 0.258 for all bond funds.¹⁰⁸ The results align for the latter's IG lagged flow coefficient, 0.362, which is similar to the fund styles that primarily invest in IG.¹⁰⁹

The **volatility** factor loadings vary across fund styles with meaningful implications for the net fund flows. First, the HY and short-term bond fund volatility coefficients are highly significant and negative at -0.519 and -0.379, respectively. On the one hand, this makes them much larger in absolute value than the volatility coefficients Chen and Qin determine: -0.134 for all, 0.000 for IG and -0.212 for HY bond funds.¹¹⁰ On the other hand, Sirri and Tufano find significant volatility coefficients of -1.043 and -1.068, which is more than twice as big.¹¹¹ Still, their magnitude relative to the average net fund flows of 0.469% makes them economically significant. This suggests that investors sell these funds shares during periods of high return volatility which could indicate economic turmoil. They are likely motivated by other reasons: HY investors potentially sell their shares with high credit, liquidity, and economic risk to reduce their exposure and minimize losses, engaging in typical flight-to-safety behavior. This potentially also applies to long-term bond fund flows with a marginally significant but large -0.585 coefficient: Long-term bond investments suffer especially during periods of changing market conditions and subsequent return volatilities and hence receive less fund flows and are sold. Therefore, their investors engage in flight-to-safety away from interest rate and economic risk. In contrast, short-term investors probably sell their low-risk shares in order to generate liquidity in a reverse flight-to-liquidity pattern. Nonetheless, the volatility has little impact on fund flows for intermediate core, intermediate core-plus, multisector, bank loan and nontraditional bond funds as the coefficients are insignificant and near zero. Chen and Qin validate this finding as their volatility coefficients are also insignificant.¹¹²

Finally, the **fund size** has no significant impact on most fund styles as the coefficients are very small and not significant. Only the corporate bond fund category has a highly significant, positive association of 0.090, indicating that these investors place minimal importance on the fund's size when allocating their capital. The standard literature is divided on the influence the fund size on the net fund flows: Chen and Qin, Goldstein et al. and Rakowski and Wang determine the coefficient to range between 0.000 and 0.001.¹¹³ Sirri and Tufano show the coefficient to be slightly negative at -0.048.¹¹⁴ Neither coefficient is large enough to be economically significant, even if it were statistically significant. Hence, the fund size is not a significant consideration for

bond fund investors.

5.3. COVID-19 Pandemic Case Study: Regressions with a Macroeconomic Focus

The impact of the COVID-19 crisis on bond funds is clear when looking at the net new fund flows of bond mutual funds over the past two decades (see Figure 5): New bond fund flows experienced a severe drop and fell to less than -10%. To illuminate the impact of the pandemic on the fund flows, I operationalize the pandemic stages in three dummy variables: Crisis, Recovery and PostCovid. Crisis applies from February to April 2020 and represents the onset and peak of the pandemic, its impact on the financial markets and the funds' resilience to the crisis.¹¹⁵ Recovery applies from May to June 2020 and tracks the recovery of the corporate bond market.

Finally, PostCovid applies from July 2020 onward; its purpose is to indicate the medium-term impact of the pandemic. The sample period remains unchanged to guarantee a comparable level of explanatory power regarding the independent and dependent variables. The VIX index is excluded, as the market volatility is highly correlated with the Crisis indicator variable and as such, it would likely create disadvantageous multicollinearity in the model. To study the association between cash ratio, performance rank and FEDFUNDS-CHG and the pandemic stages, the estimation uses interactions between these three variables and the Crisis, Recovery and PostCovid dummies, respectively. I estimate the following regression models: The net fund flows continue as dependent variable. OLS regression 6 includes all macroeconomic and fund variables and controls (see equation 7).

$$\begin{aligned}
 Flows_{i,t} = & \alpha + \beta_1 \times cash\ ratio_{i,t-1} \\
 & + \beta_2 \times FEDFUNDS-CHG_{i,t-1} \\
 & + \beta_3 \times performance\ rank_{i,t-1} \\
 & + \beta_4 \times Crisis_t + \beta_5 \times Recovery_t \\
 & + \beta_6 \times PostCovid_t \\
 & + \beta_7 \times Crisis_t \times cash\ ratio_{i,t-1} \\
 & + \beta_8 \times Crisis_t \times FEDFUNDS-CHG_{i,t-1} \\
 & + \beta_9 \times Crisis_t \times performance\ rank_{i,t-1} \\
 & + \beta_{10} \times Recovery_t \times cash\ ratio_{i,t-1} \\
 & + \beta_{11} \times Recovery_t \times FEDFUNDS-CHG_{i,t-1} \\
 & + \beta_{12} \times Recovery_t \times performance\ rank_{i,t-1} \\
 & + \beta_{13} \times PostCovid_t \times cash\ ratio_{i,t-1} \\
 & + \beta_{14} \times PostCovid_t \times FEDFUNDS-CHG_{i,t-1} \\
 & + \beta_{15} \times PostCovid_t \times performance\ rank_{i,t-1} \\
 & + \beta_{16} \times Controls_{i,t-1} + \epsilon_{i,t}
 \end{aligned}
 \tag{7}$$

¹⁰⁸cf. Goldstein et al. (2017), p. 602, cf. Y. Chen and Qin (2017), p. 6

¹⁰⁹cf. Y. Chen and Qin (2017), p. 6

¹¹⁰cf. Y. Chen and Qin (2017), p. 6

¹¹¹cf. Sirri and Tufano (1998), p. 1599

¹¹²cf. Y. Chen and Qin (2017), p. 6, 8

¹¹³cf. Y. Chen and Qin (2017), p. 6, cf. Goldstein et al. (2017), p. 602, cf. Rakowski and Wang (2009), p. 2107

¹¹⁴cf. Sirri and Tufano (1998), p. 1599

¹¹⁵cf. Falato et al. (2021), p. 41 f.

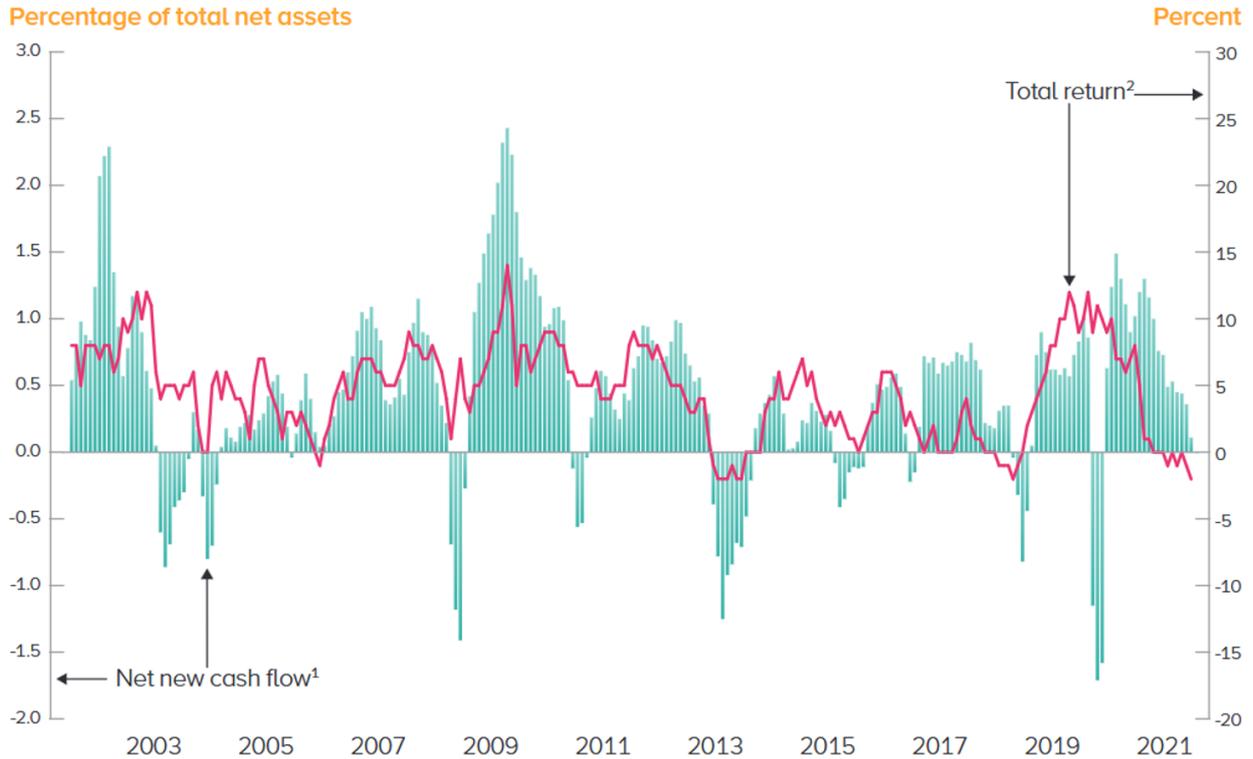


Figure 5: US bond mutual funds net new cash flows and total returns¹¹⁶

OLS regression 7 focuses on the macro variables (see equation 8).

$$\begin{aligned}
 Flows_{i,t} = & \alpha + \beta_1 \times FEDFUNDS-CHG_{i,t-1} \\
 & + \beta_2 \times Crisis_t + \beta_3 \times Recovery_t \\
 & + \beta_4 \times PostCovid_t \\
 & + \beta_5 \times Crisis_t \times FEDFUNDS-CHG_{i,t-1} \quad (8) \\
 & + \beta_6 \times Recovery_t \times FEDFUNDS-CHG_{i,t-1} \\
 & + \beta_7 \times PostCovid_t \times FEDFUNDS-CHG_{i,t-1} \\
 & + \beta_8 \times Controls_{i,t-1} + \varepsilon_{i,t}
 \end{aligned}$$

In FE regression 8, the macroeconomic variables are evaluated in a regression with Morningstar Category fixed effects (see equation 9).

$$\begin{aligned}
 Flows_{i,t} = & \alpha + \beta_1 \times FEDFUNDS-CHG_{i,t-1} \\
 & + \beta_2 \times Crisis_t + \beta_3 \times Recovery_t \\
 & + \beta_4 \times PostCovid_t \\
 & + \beta_5 \times Crisis_t \times FEDFUNDS-CHG_{i,t-1} \quad (9) \\
 & + \beta_6 \times Recovery_t \times FEDFUNDS-CHG_{i,t-1} \\
 & + \beta_7 \times PostCovid_t \times FEDFUNDS-CHG_{i,t-1} \\
 & + \beta_8 \times Controls_{i,t-1} + \eta_i + \varepsilon_{i,t}
 \end{aligned}$$

Similarly to regressions 1-5, Breusch-Pagan tests confirm heteroskedasticity in regressions 6-8 (see Table 7) and Newey-West standard errors are calculated. Table 8 presents the regression results for regression 6-8.

Table 7: Breusch-Pagan test results for the regression models 6-8

Model	Statistic	p-value
Regression 6	3981.882***	< 2.2e ⁻¹⁶
Regression 7	1129.492***	< 2.2e ⁻¹⁶
Regression 8	1573.222***	< 2.2e ⁻¹⁶

This table shows the Breusch-Pagan test results for regressions 6-8. Stars signify statistical significance: *** p < 0.001, ** p < 0.01, * < 0.05, respectively.

The COVID-19 pandemic was a notable global stressor that affected the real economy with likely lasting impacts. For this reason, I expect the dummy regression coefficients and interaction effects to be highly significant. First, the Crisis coefficient should be negative and have a high magnitude depending on the shock on, and subsequent outflows of, the corporate bond market. Second, the Recovery dummy signals the recovery of the corporate bond market, investor trust and outlook thanks to the Fed policy interventions effectively calming the markets through the interest rate monetary

¹¹⁶cp. Investment Company Institute (2022), p. 56

policy channel.¹¹⁷ Therefore, the correlation should have a smaller absolute value than the Crisis's by a wide margin, being either significantly less negative or even positive. Third, the PostCovid coefficients could indicate that markets are still actively recovering after the crisis, have largely recovered or carry a long-term impact. It would depend on their size and sign.

The correlation between the Crisis indicator variables and the fund flows is noticeably strong, i.e., highly significant with large coefficients around -2.5. This confirms that the flows decreased starkly during the crisis by up to -2.659 with style FE. These results are both larger and smaller than other studies as Falato et al. estimate the Crisis coefficient to be -0.29 and Kuong et al. determine the correlation to fund outflows to be 4.247 – still, both find that the pandemic crisis had materially decreased bond fund flows.¹¹⁸ Overall, the Crisis coefficients still align with both studies' conclusions: Bond funds experienced large outflows during the pandemic months which were driven by liquidity demand shocks.

Table 8 presents the OLS and style FE regressions for models 6 – 8. The Recovery coefficients are initially negative, but positive when focusing on macro variables and never significant. They have no clear trends since the coefficients are not significant and switch from negative to positive. This would suggest that there was no substantial difference between the fund flow dynamics before the crisis and during the recovery months, indicating a stabilization of investor behavior. A possible contributor could be the PMCCF and SMCCF policy interventions which reassured investors.¹¹⁹ With such decisive actions, investors gained confidence, and the corresponding outflows reversed.¹²⁰ The PostCovid factor loadings are initially positive and significant but lose their significance when focusing on macro variables or controlling for fund style fixed effects. Their insignificance suggests there is no major difference between fund flows before and after COVID-19. This result would corroborate the interpretation that the Fed's credit facilities calmed investors long-term, and that no negative effects remain, resonating with Falato et al., who show that investor flow dynamics normalized relatively quickly after the Fed's interventions.

Overall, the net fund flows' evolution over the Crisis, Recovery and PostCovid period mirrors the evolution of transaction costs in bond markets over the course of the pandemic, as O'Hara and Zhou show: first, the liquidity crisis at the start and height of the pandemic caused exorbitant transaction costs, then the Fed intervened with their credit facility policy and offered a liquidity backstop, leading to a quick resolution and normalized markets.¹²¹

Additionally, Falato et al. also find that COVID-related outflows were most extreme during the height of the crisis, halved in the weeks after the first Fed PMCCF and SMCCF

announcement and normalized to pre-crisis levels following the second Fed announcement which communicated the expansion to \$75 billion.¹²² These parallel behaviors support the validity of the findings of this thesis. Considering the interaction terms allows for a more differentiated interpretation of the effective federal funds rate and rate change. The FEDFUNDS-CHG interaction terms with the pandemic dummies behave as follows: the interaction between FEDFUNDS-CHG and Crisis has exceedingly negative and highly significant factor loadings in regressions 6 – 8. In contrast, FEDFUNDS-CHG and Recovery have a highly significant positive association in regressions 7 and 8. Lastly, the FEDFUNDS-CHG and PostCovid coefficients are not significant, negative and highly changeable across regressions: -0.607, 2.962 and 2.951. Hence, a one-unit increase in FEDFUNDS-CHG during the crisis leads to a decrease of at least -5.63 (= -0.813 – 4.823) in fund flows across fund styles. This finding is extremely economically significant, as this encompasses more than 45% of the net flow range (P5-P50). Next, the FEDFUNDS-CHG coefficient is 0.042 (= -1.592 + 1.634) and highly significant during Recovery.

Finally, the coefficient is 1.359 (= -1.592 + 2.951) and not significant in the months after the crisis. That shows that increases in the effective federal fund rate changes were correlated with a significant decrease in fund flows during the crisis and slight increases during the Recovery phase, possibly because the Fed rate policy actions restored investor trust. Post pandemic, the PostCovid and FEDFUNDS-CHG correlation has become relatively large. However, since it is, first, not statistically significant, and second, exceeded by its Newey-West standard error, the most that can be concluded is: there is no substantial difference in the correlation between FEDFUNDS-CHG and fund flows before the COVID-19 crisis and in the years that followed. The individual effective federal funds rate change coefficients are highly significant and negative. The FEDFUNDS factor loadings are smaller in size but still highly significant. The DEF and TB3MS factors have a highly significant, positive correlation to flows. This former's correlation is validated while the latter's is contradicted by literature.¹²³ The OPTION coefficients are slightly below zero and none are significant. This finding is not supported by Chen and Qin who estimate the OPTION factor to range between 0.028 and 0.352.¹²⁴ The adjusted R² is 0.207 in regression 6, which includes all variables, and falls drastically with the macro focus. Similarly to coefficients of determination in regressions 1-4, roughly 21% of the flow variation is explained by all variables and less than 3% by macro variables only. Still, the regression 6 adjusted R² is validated by Chen and Qin.¹²⁵

¹²²cf. Falato et al. (2021), p. 43, cf. Board of Governors of the Federal Reserve System (2020c), URL see References

¹²³cf. Y. Chen and Qin (2017), p. 7

¹²⁴cf. Y. Chen and Qin (2017), p. 8

¹²⁵cf. Y. Chen and Qin (2017), p. 6, 8

¹¹⁷cf. Falato et al. (2021), p. 37

¹¹⁸cf. Falato et al. (2021), p. 41, cf. Kuong et al. (2024), 6

¹¹⁹cf. O'Hara and Zhou (2021), p. 47

¹²⁰cf. Falato et al. (2021), p. 37

¹²¹cf. O'Hara and Zhou (2021), p. 66

Table 8: OLS and style FE regression 6-8

	Dependent variable: flows		
	All vars. OLS <i>Newey-West</i> (6)	Macro. vars. OLS <i>Newey-West</i> (7)	Macro vars. Style FE <i>Newey-West</i> (8)
volatility	0.070*** (0.016) t = 4.417		
log(fund age)	-0.397*** (0.019) t = -20.361		
past flows	0.378*** (0.008) t = 47.430		
log(TNA)	0.004 (0.009) t = 0.422		
expense ratio	-0.047 (0.049) t = -0.964		
performance rank	1.057*** (0.049) t = 21.442		
cash ratio	0.010*** (0.002) t = 4.797		
BOND	0.237*** (0.013) t = 18.812	0.320*** (0.014) t = 22.399	0.321*** (0.014) t = 22.607
FEDFUNDS	-0.375*** (0.073) t = -5.134	-0.497*** (0.109) t = -4.571	-0.493*** (0.108) t = -4.553
FEDFUNDS-CHG	-0.813*** (0.108) t = -7.523	-1.570*** (0.148) t = -10.578	-1.592*** (0.148) t = -10.741
STK	-0.016*** (0.005) t = -3.418	-0.028*** (0.005) t = -5.194	-0.029*** (0.005) t = -5.405
TB3MS	0.542*** (0.081) t = 6.686	0.722*** (0.120) t = 6.017	0.726*** (0.120) t = 6.074
DEF	0.036*** (0.008) t = 4.588	0.139*** (0.010) t = 13.849	0.141*** (0.010) t = 14.080

This table reports the fund flow regression results. Column (6) shows the OLS regression results for all variables. Column (7) shows the OLS regression for macro independent variables. Column (8) shows the FE fund style regression for macro variables. "Style FE" denotes style fixed effects. The Newey-West standard errors (in parentheses) are below. Stars show statistical significance: *** p < 0.001, ** p < 0.01, * < 0.05.

Table 8 — continued

OPTION	-0.046 (0.031) t = -1.482	-0.029 (0.036) t = -0.799	-0.029 (0.036) t = -0.792
Crisis	-2.659*** (0.278) t = -9.555	-2.482*** (0.150) t = -16.591	-2.516*** (0.149) t = -16.847
Recovery	-0.180 (0.296) t = -0.608	0.336 (0.192) t = 1.751	0.310 (0.191) t = 1.623
PostCovid	0.221* (0.088) t = 2.513	0.151* (0.070) t = 2.140	0.127 (0.070) t = 1.816
cash ratio × Crisis	-0.030* (0.013) t = -2.256		
performance rank × Crisis	0.642 (0.433) t = 1.484		
FEDFUNDS-CHG × Crisis	-4.823*** (0.344) t = -14.012	-3.722*** (0.315) t = -11.822	-3.721*** (0.315) t = -11.827
cash ratio × Recovery	-0.012 (0.012) t = -0.983		
performance rank × Recovery	1.048* (0.426) t = 2.458		
FEDFUNDS-CHG × Recovery	0.414 (0.464) t = 0.891	1.621*** (0.405) t = 4.005	1.634*** (0.404) t = 4.039
cash ratio × PostCovid	-0.008 (0.006) t = -1.335		
performance rank × PostCovid	-0.098 (0.134) t = -0.735		
FEDFUNDS-CHG × PostCovid	-0.607 (2.668) t = -0.228	2.962 (3.072) t = 0.964	2.951 (3.066) t = 0.962
Intercept	0.431** (0.164) t = 2.634	0.306*** (0.034) t = 8.891	0.329 (0.433) t = 0.760
Observations	87,160	87,160	87,160
R ²	0.207	0.022	0.026
Adjusted R ²	0.206	0.022	0.026

6. Conclusion

The aim of this thesis was to investigate the influence of fund-specific and macro variables on CBMF flows. This first required an analysis of the systematic fund flow dynamics with all variables, then with the macro and fund variables. A fund style split sample analysis was performed to inspect the individual fund flow determinants for each Morningstar Category. Finally, the impact of COVID-19 on fund flows was explored based on a dedicated case study. These made it possible to confirm well-studied correlations in one comprehensive analysis, expanding the literature by testing less-studied macro determinants, i.e., the effective federal funds rate and its change, and conducting a different pandemic case study.

In the analysis of the systematic fund flow dynamics, the effective federal funds rate showed a statistically and economically significant negative correlation to fund flows across all regressions, showing that higher rates deter investor flows due to increased liquidity costs. The corresponding rate changes, however, exhibited an insignificant relationship, indicating that the absolute level, not the changes, primarily influence fund flows. Among the fund variables, the performance rank and past flows emerged as the strongest drivers of the fund flows, highlighting strong performance chasing and flow persistence with coefficients exceeding 1. The fund style split sample analysis differentiated these results according to Morningstar Category and highlighted the different investor profiles based on investment objectives and risk tolerance. Again, performance rank was the dominant determinant across nearly all fund styles, reflecting performance-chasing behavior. Conversely, the cash ratio had minimal significance except for bank loan and high-yield funds, implying a higher responsiveness to default and liquidity risk.

Finally, the COVID-19 pandemic case study uncovered that bond mutual funds experienced severe outflows during the crisis months, consistent with large liquidity demands due to sudden investor withdrawals. The recovery period was marked by a quick fund flow stabilization indicated by materially smaller and often statistically insignificant Recovery correlations, likely due to the Fed's interventionist PM-CCF and SMCCF announcements and implementations. This development continued into the Post-Covid phase, which showed normalized behavior with no substantial fund flow difference compared to the pre-crisis period. The pandemic stage interaction terms with the changes in the effective federal funds rate revealed a heightened sensitivity during the crisis. This implies that any increases in the effective federal funds rate discourage investors and lead to more outflows. However, similar to the evolution of the pandemic, these effects dissipated during the Recovery and Post-Covid period, potentially indicating restored investor confidence and the stabilizing effect of Fed interventions.

The insignificance of the cash ratio was surprising, raising the question of if and how funds can pacify investors during financial crises. After all, as long as CBMFs must perform critical liquidity transformations to satisfy daily in-

vestor withdrawal rights, they will remain financially fragile. Since they continue to grow in size and in importance as non-bank financial intermediaries, their vulnerability remains a macro concern. Considering the macro factors, the significance of the VIX and BOND factor highlights the role of the macro market condition for fund flow dynamics; this thesis supports adding the effective federal funds rate as an important determinant. A principal limitation is the inability to model concrete key dates as I used monthly fund information. Therefore, the explanatory power of my pandemic dummies and their interaction terms is constrained, and significant fund flow movements potentially remained unobserved. Next, testing and comparing multiple liquidity measures as complementary analyses would have increased the robustness of the results, which could have been especially valuable for the analysis of the cash ratio. Indeed, the fund-level liquidity could have been calculated as bid-ask spread or Roll measure while the aggregate liquidity could be estimated using the TED factor or DFL factor.¹²⁶ Finally, future research includes extending the literature on macro determinants by testing other macro determinants and other countries. Also, green funds could be explored as Fatica et al. find that green bonds experienced lower sales during the pandemic.¹²⁷

References

- Allen, F., & Gale, D. (2004). Financial Fragility, Liquidity, and Asset Prices. *Journal of the European Economic Association*, 2(6), 1015–1048.
- Baer, J. (2020). The Day Coronavirus Nearly Broke the Financial Markets. Retrieved January 4, 2025, from <https://www.wsj.com/articles/the-day-coronavirus-nearly-broke-the-financial-markets-11589982288>
- Bao, J., Pan, J. U. N., & Wang, J. (2011). The Illiquidity of Corporate Bonds. *The Journal of Finance*, 66(3), 911–946.
- Beber, A., Brandt, M. W., & Kavajecz, K. A. (2009). Flight-to-Quality or Flight-to-Liquidity? Evidence from the Euro-Area Bond Market. *Review of Financial Studies*, 22(3), 925–957.
- Bergstresser, D., & Poterba, J. (2002). Do after-tax returns affect mutual fund inflows? *Journal of Financial Economics*, 63(3), 381–414.
- Bernanke, B. S., & Gertler, M. (1995). Inside the black box: the credit channel of monetary policy transmission. *Journal of Economic perspectives*, 9(4), 27–48.
- Board of Governors of the Federal Reserve System. (2020a). Federal Reserve Issues FOMC Statement. Retrieved January 5, 2025, from <https://www.federalreserve.gov/newsevents/pressreleases/monetary20200315a.htm>
- Board of Governors of the Federal Reserve System. (2020b). Primary Market Corporate Credit Facility. Retrieved January 5, 2025, from <https://www.federalreserve.gov/newsevents/pressreleases/files/monetary20200323b1.pdf>
- Board of Governors of the Federal Reserve System. (2020c). Secondary Market Corporate Credit Facility. Retrieved January 5, 2025, from <https://www.federalreserve.gov/newsevents/pressreleases/files/monetary20200409a5.pdf>
- Board of Governors of the Federal Reserve System. (2020d). Secondary Market Corporate Credit Facility. Retrieved January 5, 2025, from <https://www.federalreserve.gov/newsevents/pressreleases/files/monetary20200323b2.pdf>
- Board of Governors of the Federal Reserve System. (2025). About the Federal Reserve. Retrieved January 5, 2025, from <https://www.federalreserve.gov/aboutthefed.htm>

¹²⁶cf. Falato et al. (2021), p. 43, cf. Goldstein et al. (2017), p. 604

¹²⁷cf. Fatica and Panzica (2024), p. 18

- Boyarchenko, N., Kovner, A., & Shachar, O. (2022). It's what you say and what you buy: A holistic evaluation of the corporate credit facilities. *Journal of Financial Economics*, 144(3), 695–731.
- Bush, S. (2019a). Do You Need a Multisector Bond Fund? Retrieved January 5, 2025, from <https://www.morningstar.com/bonds/do-you-need-multisector-bond-fund-2>
- Bush, S. (2019b). Introducing Two New Morningstar Bond Categories. Retrieved January 5, 2025, from <https://www.morningstar.com/funds/introducing-two-new-morningstar-bond-categories>
- Chen, Q., Goldstein, I., & Jiang, W. (2010). Payoff complementarities and financial fragility: Evidence from mutual fund outflows. *Journal of Financial Economics*, 97(2), 239–262.
- Chen, Y., & Qin, N. (2017). The Behavior of Investor Flows in Corporate Bond Mutual Funds. *Management Science*, 63(5), 1365–1381.
- Chevalier, J., & Ellison, G. (1997). Risk Taking by Mutual Funds as a Response to Incentives. *Journal of Political Economy*, 105(6), 1167–1200.
- Choi, J., Hoseinzade, S., Shin, S. S., & Tehrani, H. (2020). Corporate bond mutual funds and asset fire sales. *Journal of Financial Economics*, 138(2), 432–457.
- Choi, J., & Kronlund, M. (2018). Reaching for Yield in Corporate Bond Mutual Funds. *Review of Financial Studies*, 31(5), 1930–1965.
- Edelen, R. M., & Warner, J. B. (2001). Aggregate price effects of institutional trading: a study of mutual fund flow and market returns. *Journal of Financial Economics*, 59(2), 195–220.
- Elton, E. J., Gruber, M. J., & Blake, C. R. (1995). Fundamental Economic Variables, Expected Returns, and Bond Fund Performance. *The Journal of Finance*, 50(4), 1229–1256.
- European Central Bank. (2025). The Transmission Mechanism of Monetary Policy. Retrieved January 5, 2025, from <https://www.ecb.europa.eu/mopo/intro/transmission/html/index.en.html>
- Falato, A., Goldstein, I., & Hortaçsu, A. (2021). Financial fragility in the COVID-19 crisis: The case of investment funds in corporate bond markets. *Journal of Monetary Economics*, 123, 35–52.
- Fatica, S., & Panzica, R. (2024). Sustainable investing in times of crisis: Evidence from bond holdings and the COVID-19 pandemic. *Journal of Banking & Finance*, 166, 1–23.
- Federal Reserve Bank of New York. (2020). Corporate Credit Facility: Frequently Asked Questions. Retrieved January 4, 2025, from <https://www.newyorkfed.org/markets/primary-and-secondary-market-faq/archive/corporate-credit-facility-faq-201204>
- Federal Reserve Bank of New York. (2025). Effective Federal Funds Rate. Retrieved January 1, 2025, from <https://www.newyorkfed.org/markets/reference-rates/effr>
- Financial Stability Board. (2024a). Global Monitoring Report on Non-Bank Financial Intermediation: Data. Retrieved January 4, 2025, from <https://www.fsb.org/work-of-the-fsb/financial-innovation-and-structural-change/non-bank-financial-intermediation/global-nbfi-monitoring-report-data/>
- Financial Stability Board. (2024b). Non-bank financial intermediation. Retrieved January 5, 2025, from <https://www.fsb.org/work-of-the-fsb/financial-innovation-and-structural-change/non-bank-financial-intermediation/>
- Goldstein, I., Jiang, H., & Ng, D. T. (2017). Investor flows and fragility in corporate bond funds. *Journal of Financial Economics*, 126(3), 592–613.
- Hirshleifer, D. (2015). Behavioral Finance. *Annual Review of Financial Economics*, 7(1), 133–159.
- Idzelis, C. (2020). The Corporate Bond Market Is 'Basically Broken,' Bank of America Says. *Institutional Investor*. Retrieved January 5, 2025, from <https://www.institutionalinvestor.com/article/2b5wze3jqc khuckroy7sw/portfolio/the-corporate-bond-market-is-basically-broken-bank-of-america-says>
- Investment Company Institute. (2022). *2022 Investment Company Fact Book: A Review of Trends and Activities in the Investment Company Industry*.
- Investopedia. (2025). Heteroskedasticity. Retrieved January 5, 2025, from <https://www.investopedia.com/terms/h/heteroskedasticity.asp>
- Jiang, H., Li, D., & Wang, A. (2021). Dynamic Liquidity Management by Corporate Bond Mutual Funds. *Journal of Financial and Quantitative Analysis*, 56(5), 1622–1652.
- Kramer, M. J. (2024). Over-the-Counter Market. Retrieved January 4, 2025, from <https://www.investopedia.com/terms/o/over-the-counter-market.asp>
- Kuong, J. C., O'Donovan, J., & Zhang, J. (2024). Monetary policy and fragility in corporate bond mutual funds. *Journal of Financial Economics*, 161, 103931.
- Lagunoff, R., & Schreft, S. L. (2001). A Model of Financial Fragility. *Journal of Economic Theory*, 99(1-2), 220–264.
- Longstaff, F. A. (2004). The Flight to Liquidity Premium in U.S. Treasury Bond Prices. *The Journal of Business*, 77(3), 511–526.
- Ma, Y., Xiao, K., & Zeng, Y. (2022). Mutual Fund Liquidity Transformation and Reverse Flight to Liquidity. *Review of Financial Studies*, 35(10), 4674–4711.
- Mishkin, F. S. (1995). Symposium on the Monetary Transmission Mechanism. *Journal of Economic Perspectives*, 9(4), 3–10.
- Morningstar. (2025a). Corporate Bond Funds. Retrieved January 5, 2025, from <https://www.morningstar.com/best-investments/corporate-bond-funds>
- Morningstar. (2025b). High-Yield Bond Funds. Retrieved January 5, 2025, from <https://www.morningstar.com/best-investments/high-yield-bond-funds>
- Morningstar Office. (2023a). Bank Loan Category. Retrieved January 5, 2025, from https://awgmain.morningstar.com/webhelp/glossary_definitions/categories/Bank_Loan_Category.htm
- Morningstar Office. (2023b). Long-Term Bond Category. Retrieved January 5, 2025, from https://awgmain.morningstar.com/webhelp/glossary_definitions/categories/Long_Term_Bond_Category.html
- Morningstar Office. (2023c). Multisector Bond Category. Retrieved January 5, 2025, from https://awgmain.morningstar.com/webhelp/glossary_definitions/categories/Multisector_Bond_Category.html
- Morningstar Office. (2023d). Nontraditional Bond Category. Retrieved January 5, 2025, from https://awgmain.morningstar.com/webhelp/glossary_definitions/categories/Nontraditional_Bond_Category.html
- Morningstar Office. (2023e). Short-Term Bond Category. Retrieved January 5, 2025, from https://awgmain.morningstar.com/webhelp/glossary_definitions/categories/Short_Term_Bond_Category.html
- O'Hara, M., & Zhou, X. (2021). Anatomy of a liquidity crisis: Corporate bonds in the COVID-19 crisis. *Journal of Financial Economics*, 142(1), 46–68.
- O'Hara, M., & Zhou, X. (2023). Things Fall Apart: Fixed Income Markets in the COVID-19 Crisis. *Annual Review of Financial Economics*, 15(1), 55–68.
- Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2015). Scale and skill in active management. *Journal of Financial Economics*, 116(1), 23–45.
- PIMCO. (n.d.). Everything You Need to Know About Bonds. Retrieved February 1, 2025, from <https://www.pimco.com/eu/en/resources/education/everything-you-need-to-know-about-bonds>
- Rakowski, D., & Wang, X. (2009). The dynamics of short-term mutual fund flows and returns: A time-series and cross-sectional investigation. *Journal of Banking & Finance*, 33(11), 2102–2109.
- Sirri, E. R., & Tufano, P. (1998). Costly Search and Mutual Fund Flows. *The Journal of Finance*, 53(5), 1589–1622.
- U.S. Securities and Exchange Commission. (2013a). What Are Corporate Bonds? Retrieved January 4, 2025, from <https://www.investor.gov/introduction-investing/general-resources/news-alerts/alerts-bulletins/investor-bulletins/what-are>
- U.S. Securities and Exchange Commission. (2013b). What Are High-yield Corporate Bonds? Retrieved January 4, 2025, from <https://www.investor.gov/introduction-investing/general-resources/news-alerts/alerts-bulletins/investor-bulletins/what-are-0>



The Influence of Leadership Style on the Acceptance of Generative AI in the Workplace - The Role of Organizational Commitment, Job Insecurity and Interaction Frequency

Felix Yumuşak

Freie Universität Berlin

Abstract

This thesis investigates the influence transformational and transactional leadership styles have on the acceptance of generative AI in the workplace. To account for other factors, I also examine the mediating effect of organizational commitment and the fear of job loss as well as the moderating effect of interaction frequency with the supervisor. Using a sample of 220 full time working participants, I find that technology acceptance does not significantly differ between transformational and transactional leadership. However, the results show that the fear of job loss significantly mediates the relationship between leadership and technology acceptance. While organizational commitment does not mediate the relationship, it does significantly influence technology acceptance. My research extends research incorporating threat rigidity theory by showing that transformational leadership reduces AI-related fear of job loss. Therefore, it highlights the importance of considering additional factors such as prior experience and openness to innovations in AI acceptance. The findings suggest that managers should adopt transformational leadership to decrease the fear of job loss and enhance organizational commitment to effectively increase AI acceptance in the workplace.

Keywords: AI acceptance; fear of job loss; Gen AI; leadership style; technology acceptance

1. Introduction

In recent years, the rapid advancement of generative artificial intelligence (AI) technologies has opened new opportunities for innovation and efficiency in the workplace. AI is becoming a central force in society, transforming industries and societal structures while raising questions about ethics, privacy, and social inequality (Khogali & Mekid, 2023). Generative AI refers to AI systems that can generate unique outputs, such as text, images, or code in a conversational manner. These AI applications (e.g. ChatGPT, Claude, Gemini) are trained on extensive datasets, enabling them to engage users in conversational interactions that yield human-like responses (Cascella et al., 2023; Fui-Hoon Nah et al., 2023). The most popular AI ‘ChatGPT’ can significantly increase productivity in professional writing tasks. It can reduce the average time taken by 40% and enhance output quality by 18% (Noy & Zhang, 2023). A study by KPMG (2024) found that

67% of managers expect generative AI to increase sales and automation and 65% expect efficiency gains to reduce costs. However, 37% expect job losses and employee acceptance issues. According to Deloitte (2023), 43% of employees say that they are worried about losing their jobs due to the increased use of AI programs over the next five years. These statistics show the perceived potential of generative AI to drive progress but also disrupt and change the labor market.

Leadership plays a crucial role in facilitating technology acceptance (Neufeld et al., 2007). Although the influence of leadership style on the acceptance of new technologies has been studied extensively (Aziz et al., 2020; Molino et al., 2021; Schepers et al., 2005), these studies have primarily focused on technologies such as software, web-based applications, and robots. The influence of leadership style on the acceptance of generative AI in the workplace remains underexplored, as existing studies do not address this specific

context. The unique characteristics of AI, such as its creativity, conversational nature, and ability to learn, present distinct challenges. These challenges and opportunities differ significantly from those associated with other technologies. This thesis contains an exploration of the influence of leadership style on the acceptance of generative AI in the workplace. Understanding how different leadership approaches impact AI acceptance can provide valuable insights for managers who want to foster a culture of technology acceptance.

The specific challenges and opportunities presented by AI technologies have not been addressed in studies of organizational commitment and technology acceptance. New technologies like AI can also spark the fear of job losses due to automation (McClure, 2018; Ore & Sposato, 2022). While research shows a positive relationship between transformational leadership and organizational commitment, and the influence of organizational commitment on technology acceptance or change has been examined, AI technologies as a unique and innovative technology remain a gap in the literature. Avolio et al. (2004) demonstrated the relationship between transformational leadership and organizational commitment, and Vella et al. (2013) and Iverson (1996) explored the influence of organizational commitment on technology acceptance and organizational change, but they did not focus on the opportunities and challenges generative AI technologies present.

The insights obtained in this thesis advances the understanding of generative AI acceptance in the workplace by an examination of the differential influence of transformational and transactional leadership styles. I argue that transformational leadership significantly enhances AI acceptance relative to transactional leadership, primarily through its positive influence on perceived usefulness. My analysis examines two key mediating factors: organizational commitment, which I propose increases technology acceptance, and fear of job loss, which I propose decreases it. I also examine how the frequency of employee-manager interactions moderates the relationship between transformational leadership and AI acceptance. I propose that employees are more likely to accept generative AI technologies when they are led by transformational leaders who effectively communicate the benefits of AI, foster organizational commitment, and mitigate fears of job displacement. This effect is particularly pronounced in contexts where there is frequent interaction between leaders and employees, allowing for greater influence on perceptions and attitudes toward new technologies.

Building on this framework, the goal of this thesis is to contribute to the academic discourse on leadership and technology acceptance by extending theoretical concepts into the emerging context of generative AI in the workplace. By empirically examining how transformational and transactional leadership styles influence AI acceptance, this study addresses a significant gap in the literature. Furthermore, by introducing organizational commitment and fear of job loss as key mediating variables, the research offers new insights into the mechanisms through which leaders can facilitate AI acceptance. These findings not only enhance our theoretical

understanding but also provide a valuable foundation for future research to explore the relationship between leadership, organizational dynamics, and technology acceptance in rapidly evolving environments.

This thesis is structured as follows: First, I review relevant psychological and organizational theories and research on leadership style and develop four hypotheses. This is followed by presenting the chosen method. I then present my empirical results. Following, I discuss my findings, describe my theoretical contributions and managerial implications and state the limitations and future research. I then conclude the thesis with a short summary.

2. Theory and Hypotheses

In this section, I present the relevant leadership theories, psychological frameworks, and organizational concepts that form the basis for the subsequent development of my hypotheses.

2.1. Leadership Theories

2.1.1. Transformational Leadership

Transformational leadership, initially defined by Burns (1978), has been extensively studied and developed over the years. Burns proposed that transformational leaders motivate and inspire followers to surpass their immediate self-interests by identifying with a broader vision, whereas transactional leaders focus on exchanges and rewards to obtain cooperation (Judge & Bono, 2000). This distinction established the foundation for further investigation into the nature and implications of transformational leadership. One of the most comprehensive expansions on Burns' original concept was provided by Bass (1985), who developed a multidimensional approach for the understanding of transformational leadership. According to Bass, see Figure 1, transformational leadership consists of four key dimensions: idealized influence, inspirational motivation, intellectual stimulation, and individualized consideration (Bass, 1999). These dimensions explain how transformational leaders act as role models, inspire with a clear vision, stimulate intellectual challenges, and attend to individual follower needs.

Building on this framework, Bass further elaborated on these dimensions, providing detailed insights into their practical application. Leaders who demonstrate *idealized influence* and *inspirational leadership* visualize and describe a promising and desirable future and clearly articulate the path to get there. They set a strong example for others to follow and demonstrate determination and confidence. When leaders engage in *intellectual stimulation*, they encourage followers (or employees) to be more innovative and to enhance their creativity. *Individualized consideration* is displayed when leaders focus on the followers' developmental needs and provide support and coaching to encourage their growth. As part of this consideration, tasks are delegated as opportunities for personal and professional growth (Bass, 1999).

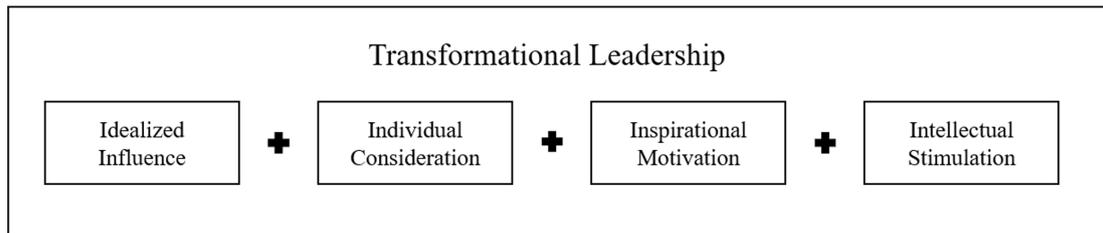


Figure 1: Elements of Transformational Leadership, adapted from Bass and Avolio (1990, p. 231)

Bass (1999) further states that idealized influence and inspirational motivation often merge into what is commonly recognized as charisma, a quality that makes leaders appear visionary and appealing to followers. Intellectual stimulation promotes creativity by questioning assumptions and encouraging innovative thinking, while individualized consideration focuses on the personal development of followers, which is a more in-depth approach compared to traditional leadership models (Judge & Bono, 2000).

Beyond individual and immediate organizational impact, transformational leadership also seeks to incite changes in organizational culture and values (Pawar & Eastman, 1997). Thus transformational leadership can be crucial in guiding organizations through periods of significant change (Eisenbach et al., 1999; Faupel & Süß, 2019; Peng et al., 2021). Herold et al. (2008) demonstrated that transformational leaders play a key role in guiding organizations through change. They enable the alignment of employees with new visions and provide essential support. The ability to adapt is important when markets are frequently disrupted by new technologies or regulations. However, the effectiveness of Transformational leadership can vary significantly depending on organizational, social, and cultural contexts (Pawar & Eastman, 1997). Although the fundamental attributes of transformational leadership remain constant, their implementation must be adapted to align with the distinctive characteristics of each context (Pawar & Eastman, 1997).

Research has also shown that transformational leadership is not only effective but also more satisfying to followers compared to transactional approaches. Leaders who exhibit transformational qualities tend to be viewed as more effective and more satisfying by their followers (Bass, 1999). This effectiveness extends across different demographic groups, with studies indicating that women tend to exhibit more transformational leadership qualities than men, potentially due to overcoming biases and focusing on effective leadership behaviors (Bass, 1999).

While transformational leadership has many strengths, critical perspectives exist. van Knippenberg and Sitkin (2013) have criticized charismatic-transformational leadership research for its conceptual ambiguity, insufficient causal models, and measurement limitations, recommending a shift to more precisely defined and empirically distinct leadership aspects. Furthermore, ethical concerns arise when transformational leaders exploit their followers for personal gain or non-beneficial goals. Tourish and Pinnington (2002) argued

that such leaders may abuse their power, leading to a cult-like following that undermines critical thinking and autonomy, ultimately causing harm to individuals and the organization.

2.1.2. Transactional Leadership

Transactional leadership places a strong emphasis on the exchange between leaders and followers. It is often contrasted with transformational leadership and is characterized by a focus on the fulfillment of specific tasks and responsibilities through a system of rewards and punishments (Bass, 1985). Transactional leadership defines leadership as a set of exchanges between leaders and followers. In this model, leaders administer rewards contingent on the performance of followers, and correct non-compliance or inadequate performance through penalties or corrective action. This approach, see Figure 2, is supported by two primary mechanisms: management by exception and contingent rewards (Bass, 1985).

Bass (1985, p. 122) states that “both contingent reward and contingent penalization are characteristics of transaction oriented managers because such managers unlike transforming leaders, are more concerned with efficient processes than with substantive ideas”. In a contingent reward system, the leader establishes clear expectations and goals for followers and provides specific rewards when those expectations are met. This type of reward system is based on the premise that followers are motivated to achieve desired outcomes when they are aware of the tangible benefits that will follow their performance. The rewards can include financial incentives or recognition. By aligning rewards with performance, leaders can effectively drive motivation and ensure that followers’ actions are aligned with organizational objectives (Bass, 1985).

Management by exception is a leadership style in which the leader only takes corrective action when followers deviate from established standards or when their performance falls below expectations (Bass, 1985). There are two forms of management by exception: active and passive. In the active form, leaders monitor followers’ performance closely and intervene proactively to correct mistakes or deviations (Bass, 1985). In the passive form, leaders only intervene after problems have become serious or after performance issues have been brought to their attention. This approach allows leaders to direct their attention to significant issues while maintaining a level of supervision to ensure that targets are met (Bass, 1985).

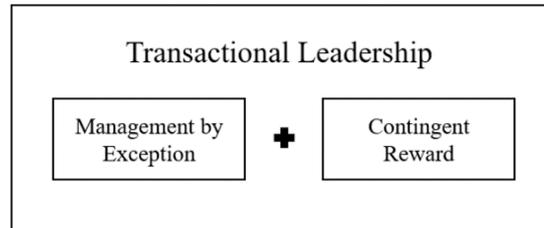


Figure 2: Elements of Transactional leadership, adapted from Bass and Avolio (1990, p. 231)

Transactional leadership is effective in fulfilling immediate goals through resource-based exchanges but lacks the capacity to inspire significant changes in followers' beliefs and values (Kuhnert & Lewis, 1987). Compared to transformational leadership, transactional leadership has been found to negatively affect employees' well-being, including higher levels of stress and lower job satisfaction. This can result in a reduction in overall productivity and an increase in turnover rates (Lyons & Schneider, 2009). Nielsen et al. (2019) demonstrated, that transformational leadership and contingent rewards (transactional leadership) are not compatible and rather contradict each other.

2.1.3. Leader Member Exchange

The Leader-Member Exchange (LMX) theory was originally developed by George Graen and William J. Haga in the 1970s (G. Graen & Cashman, 1975). It provides a robust framework for understanding the relationships between leaders and their followers in organizations. According to LMX theory, leadership is essentially an interaction process between leaders and followers that is characterized by the quality of their relationship. The central proposition of LMX is that leaders develop unique, individualized relationships with each follower, resulting in differentiated exchange relationships that can be categorized as either high or low quality exchange (G. B. Graen & Uhl-Bien, 1995). Critical perspectives argue that the potential for bias and favoritism inherent in differentiated leader-member relationships can lead to perceptions of unfairness and inequity among followers. In particular for those who are part of low-quality exchanges. Additionally, critics argue that LMX theory places too much emphasis on dyadic relationships at the expense of understanding the broader network of interactions within the organization (Schyns & Day, 2010).

2.2. Organizational Theories

2.2.1. Organizational Commitment

Mowday, Steers and Porter defined organizational commitment as "(...) the relative strength of an individual's identification with and involvement in a particular organization" (1982, p. 27). They posited that organizational commitment involves an active relationship between individuals and the organization. In order to contribute to the organization's benefit, individuals can put their goals second (Mowday et al., 1982). Building on this concept, Meyer

and Allen developed the three component model of organizational commitment. They distinct between affective, continuance and normative commitment (Meyer & Allen, 1991). Affective commitment is defined as the emotional attachment to the organization. Employees stay within the organization because they want to (Meyer & Allen, 1991). The costs of leaving the organization are associated with continuance commitment which leads to employees staying because they need to stay. Normative commitment refers to the feeling that employees have the feeling that they should stay with the organization due to e.g. rewards in advance (Meyer & Allen, 1991).

Participative leadership and leader communication has been found as antecedents of organizational commitment (Mathieu & Zajac, 1990). Several studies provided evidence for the relationship between Participative and transformational leadership and organizational commitment (Dale & Fox, 2008; Huey Yiing & Zaman Bin Ahmad, 2009; Yousef, 2000). Research has shown that transformational leadership has a stronger positive correlation with organizational commitment than transactional leadership (L. Y. Chen, 2004; Ismael et al., 2011; Limsila & Ogunlana, 2008). Erkutlu (2008) found that transformational leadership is positively correlated with organizational commitment, while management by exception negatively affects it. However, other studies found a positive relationship between transactional leadership and organizational commitment (Ramli et al., 2019). Table 1 summarizes what key studies found about the relationship of leadership style and organizational commitment.

Consequences of organizational commitment such as performance are likely moderated by financial compensation (Mathieu & Zajac, 1990). Chughtai and Zafar (2006) discovered that commitment negatively related to turnover intentions and positively linked to a self-report measure of job performance. In study of Randall (1990) it was found that while organizational commitment has some positive effects on work outcomes, these effects are generally weak and influenced by various methodological and conceptual factors. Affective and normative commitment play an important role in predicting positive change-related behaviors, while continuance commitment often correlates with mere compliance (Herscovitch & Meyer, 2002).

2.2.2. Threat Rigidity Theory

According to Staw et al. (1981) threat rigidity theory posits that when an organization perceives a threat, two pri-

Table 1: Studies of the relationship of leadership styles and organizational commitment

Study	Leadership Type	Relationship with Organizational Commitment
L. Y. Chen, 2004; Dale and Fox, 2008; Huey Yiing and Zaman Bin Ahmad, 2009; Ismael et al., 2011; Limsila and Ogunlana, 2008; Yousef, 2000	Transformational Leadership	Positive correlation
Raja and Palanichamy, 2011	Transactional Leadership	Less correlated than transformational leadership
Erkutlu, 2008	Transformational Leadership	Positive correlation
Erkutlu, 2008	Management by Exception	Negative correlation

mary effects occur. The result of a threat can be a restriction of information processing like a narrowing attention field or a constriction in control like a higher concentration of power and influence. A threat causes changes in a system's information and control processes. These changes in information processing and control reduce the diversity and flexibility of the system's behavior, leading to more rigid and less adaptive responses to the threat, see Figure 3 (Staw et al., 1981).

Structural and technological shifts are redefining work dynamics, causing significant transformations in the economy and market structures. Sanchez-Gomez et al. (2021) demonstrated that economic stress influences innovative work behaviors based on the threat rigidity theory and the conservation of resources theory. In a study from Mariano et al. (2022) it was found that older adults tend to underuse technology due to the threat of confirming ageist stereotypes. Thus (Stereotype) threat may be an important barrier to technology acceptance and usage. Employees may display threat rigidity, strictly adhering to known practices to avoid the risks associated with new technology, which stifles creativity and reduces the potential benefits of the new technology (Koberg et al., 2003). Under the influence of threat rigidity, employees may become resistant to changes (Sarkar & Osiyevskyy, 2018).

2.3. Psychological Theories

2.3.1. Conservation of Resources Theory

The conservation of resources (COR) theory was developed in 1989 by Stevan E. Hobfoll. It states that individuals are motivated to protect (conserve) their current resources and to acquire new resources. Hobfoll defines resources "(...) as those objects, personal characteristics, conditions, or energies that are valued by the individual or that serve as a means for attainment of these objects, personal characteristics, conditions, or energies." (1989, p. 516). The value of resources differs from person to person and is closely tied to individual experiences and circumstances. Individuals seek to protect their resources and feel stress when their resources are threatened or lost (Hobfoll, 1989). COR theory further states that individuals must invest resources to protect against resources loss. In the context of resource loss, the

importance of new resource gain increases. This is because when circumstances of resource loss are prevalent, the value of resource gains is enhanced. Furthermore, when resources are depleted, individuals go into a protective state in order to preserve themselves (Hobfoll et al., 2018). Employment can be considered a resource, specifically a condition according to Hobfoll.

According to COR theory, it is more stressful to lose resources than it is satisfying to gain new resources (Cacioppo & Gardner, 1999; Hobfoll, 2001). This principle suggests that negative changes (losses) at work, such as a reduction in tasks or responsibilities, have a more significant impact than positive changes, such as increased productivity due to mentioned losses. A study by Toh et al. (2023) linked COVID-19 anxiety to increased technology adoption among academics through the moderating effects outlined in COR theory.

Despite its success, COR theory has faced criticism due to the failure to consider cultural differences regarding the classification of resources. Since cultures perception of resources can substantially differ this potentially limits the cross-cultural applicability of the COR theory (Halbesleben et al., 2014; Morelli & Cunningham, 2012). Additionally, the COR theory has been criticized due to the methodical difficulty in measuring resources and quantifying their value (Ford et al., 2007).

2.3.2. Social Exchange Theory

Social Exchange Theory (SET) belongs to the most discussed paradigms for understanding behavior at work (Cropanzano & Mitchell, 2005). The basic concepts were first introduced by George C. Homans in the late 1950s. Homans proposed the idea of "social behavior as exchange" in 1958, emphasizing the psychological aspects and instrumental behavior within social exchanges (Homans, 1958). John Thibaut (1959) contributed to the SET with his concept of "social psychology of groups" by introducing the idea of interdependence, where individuals' outcomes are influenced by others in the group. Thibaut also introduced the concepts of *comparison level* and *comparison level for alternatives*, explaining how individuals evaluate the quality of their relationships and decide whether to stay or leave based on

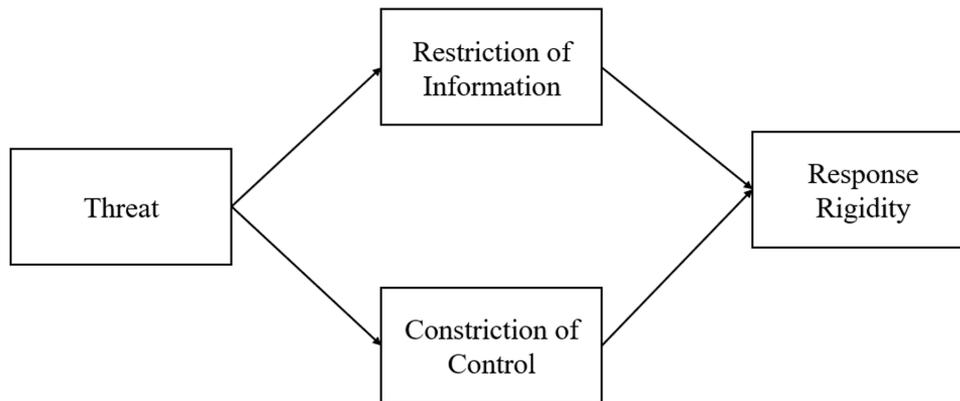


Figure 3: How threats lead to restricted information, centralized control, and rigid responses: Threat Rigidity adapted from Staw et al. (1981, p. 502)

perceived alternatives. Peter Blau (1964) expanded the theory to include concepts of power and economic orientation, emphasizing the economic dimensions and reciprocal nature of social exchange.

The key concepts of the SET is the principle of reciprocity (Cropanzano & Mitchell, 2005). Gouldner (1960) differentiated reciprocity into three distinct categories: transaction, belief, and moral norm. Transactional reciprocity refers to interdependent (both depend on each other) exchanges while reciprocity as a belief is centered around the idea of cultural orientation. Gouldner (1960) also has posited general moral norm principle of reciprocity outlines certain actions and responsibilities in exchange for benefits received.

By adding the concept of psychological transactions and emphasizing implicit (inactive) and explicit (active) exchanges, Cropanzano et al. (2017) redefined SET. Psychological exchanges encompass intangible elements such as feelings, emotions, and mental states that are not always directly observable. Unspoken or non-physical interactions that nevertheless affect the social exchange process are referred to as implicit exchange, also known as inactive exchange. Explicit exchange, or active exchange, involves direct and visible interactions, such as verbal praise, constructive feedback, or direct assistance with tasks (Cropanzano et al., 2017).

Homans (1961) associated social exchange with a rewarding or costing, more or less tangible exchange or interaction between individuals. According to Homans costs can be seen as alternative activities or opportunities. Individuals are more likely to remain in a relationship and make investments when the potential rewards outweigh the potential costs. Conversely, if the costs outweigh the rewards, individuals may attempt to alter or terminate the relationship in favor of more fulfilling alternatives (Thibaut, 1959).

2.4. Technology Acceptance

2.4.1. Theory of Reasoned Action

Fishbein and Ajzen developed the Theory of Reasoned Action (TRA) in 1975 with the objective of predicting how individuals will behave based on their pre-existing attitudes and

behavioral intentions (Fishbein & Ajzen, 1975). The behavioral intention (BI) is shaped by the attitude (A) towards the behavior and the subjective norm (SN). An individual's attitude toward a behavior is influenced by their positive or negative beliefs about it. These attitudes are the result of a lifetime of accumulated beliefs. Subjective norms are defined as an individual's perceptions of how others view a specific behavior and the perceived social pressure to either perform or abstain from that behavior (Ajzen, 1980). The behavioral intention then predicts the final, voluntary behavior. The TRA can be put into the equation:

$$BI = A + SN \quad (1)$$

It is acknowledged that external factors exist beyond the scope of the model, recognizing that the TRA is not a closed system (Ajzen, 1980). Several Studies have applied the TRA in the field of technology acceptance research (Buabeng-Andoh, 2018; Mishra et al., 2014; Yousafzai et al., 2010). While the TRA has been criticized for its simplicity and broad applicability, these same characteristics are also considered its greatest strengths, as they contribute to its generalizability across different contexts (Davis, 1985; Madden & Ajzen, 1992; Sheppard et al., 1988).

2.4.2. Theory of Planned Behavior

The Theory of Planned Behavior (TPB) is an extension of the TRA. By including behavioral control, Ajzen (1991) expanded the TRA to predict behaviors in which individuals have full volitional control. The TPB explains the behavior as predicted by the intention (I). The intention is shaped by three components. Attitude towards the behavior (A) which is influenced by behavioral beliefs, subjective norm (SN) which is influenced by normative beliefs and as a new component in the TPB the perceived behavioral control (PBC) which is influenced by control beliefs (Ajzen, 1991). PBC according to Ajzen (1991) refers to an individual's belief in their ability to perform a specific behavior, considering factors that may facilitate or hinder that performance. It

includes both the individual's confidence in their ability and the perceived ease or difficulty of the behavior. The TPB can be translated to the equation:

$$I = A + SN + PBC \quad (2)$$

The TPB has been used to explain decisions regarding the use of educational technology (Lee et al., 2010), consumer adaptation of e-services (S.-C. Chen & Li, 2010), technology adoption of mobile banking (Aboelmaged & Gebba, 2013) and the acceptance of instant messaging (Lu et al., 2009). Recent studies have applied the TPB to the adoption of AI in agriculture (Mohr & Kühn, 2021), higher education (Ivanov et al., 2024) and the workplace (Srivastava et al., 2024). Based on the TPB Ozkan and Kanat (2011) developed and validated by advanced statistical techniques an e-Government adoption model. The results demonstrated that the predictor variables of the TPB can predict e-government adoption behavior to a sufficient degree. The most significant predictor of intentions and perceived behavioral control was attitude. The second most significant predictor was perceived behavioral control. The findings of Shih and Fang (2004) indicate that intention to adopt Internet banking as a new innovation can be explained by attitude while subjective norm did not demonstrate a significant influence on intention.

While examining whether authentic leadership influences organizational change, Bakari et al. (2017) incorporated TPB to explain how leader-follower interaction influence mechanisms. They linked readiness for change and commitment to change to the TPB dimensions which their analysis supported. Gu et al. (2024) conducted a study based on the TPB that investigated the impact of empowering leadership on employee innovation behavior.

2.4.3. Technology Acceptance Model

The Technology Acceptance Model (TAM), see Figure 4., is built on two related bodies of literature, the TRA and the TPB (Canziani & MacSween, 2021).

The TAM was developed by Davis in 1989 (Davis, 1989). It has been frequently used to understand user adoption and acceptance of new technologies (McLean & Osei-Frimpong, 2019). Davis (1989) states that an individual's willingness to use technology is influenced by their attitude (A) toward it, as well as their perception of its usefulness and ease of use. Perceived usefulness (PU) refers to an individual's perception of how an information system benefits them in an organizational context. Davis et al. (1989) defined perceived ease of use (PEOU) as the user's expectation that the system will be easy to use. The behavioral intention to use (BI) is influenced by A and PU which are influenced by PEOU. Davis et al. (1989) used the following regressions to show the antecedents of BI:

$$BI = A + PU \quad (3)$$

The TAM suggests that, all else being equal, individuals form behavioral intentions based on their positive affect, as

represented by the A - BI relationship. Equation (3) describes the PU - BI relationship as well, which suggests that individuals form intentions for behaviors that they believe will improve job performance, regardless of positive or negative feelings towards the behavior itself.

In Contrast to the TRA the TAM does not include SN as a determinant of BI. SN is according to Fishbein and Ajzen (1975) the least understood aspect of the TRA. Davis et al. (1989) did not include SN in the TAM due to uncertain theoretical and psychometric status. The TAM states that A is determined by U and EOU, with relative weights evaluated using linear regression:

$$A = PU + PEOU \quad (4)$$

Davis et al. (1989) hypothesized that PU has a positive influence on A which is consistent with previous information systems research (Barrett et al., 1968; Schultz & Slevin, 1975). Davis et al. further state that PEOU is supposed to have a significant effect on A. The influence PEOU has on A can be distinguished in self-efficacy and instrumentality. Self-efficacy refers to acting independently from instrumental drivers of behavior. Instrumentality means that individuals may be contributing more to a certain target in terms of performance due to increased PEOU (Davis et al., 1989). To account for various external variables like learning based on feedback, objective design or system characteristics Davis et al. (1989) state the following equation:

$$PU = PEOU + \text{External Variables} \quad (5)$$

External variables as in equation (5) "(...) provide the bridge between the internal beliefs, attitudes and intentions represented in TAM and the various individual differences, situational constraints and managerially controllable interventions impinging on behavior" (Davis et al., 1989, p. 988). In studies various external variable have been identified such as enjoyment and computer anxiety (Abdullah et al., 2016; Castiblanco Jimenez et al., 2021), age (Burton-Jones & Hubona, 2006), experience, facilitating conditions, individual innovativeness, system quality and social norm (Castiblanco Jimenez et al., 2021).

Despite its popularity the TAM has faced criticism by several researchers. Bagozzi (2007) posits that the TAM has significant gaps between intentions and behavior, as well as between PEU, PEOU and intention. The emphasis on PU and PEOU has made it difficult to incorporate other relevant beliefs, such as trust and enjoyment, into the model (Barki & Benbasat, 2007; Malatji et al., 2020). Barki and Benbasat have criticized that the TAM neglects other important user behaviors such as reinvention, learning, and adaptation by emphasizing system use as a single, narrow behavior (Barki & Benbasat, 2007). Important aspects such as costs and structural requirements are overlooked by the TAM according to Malatji et al. (2020).

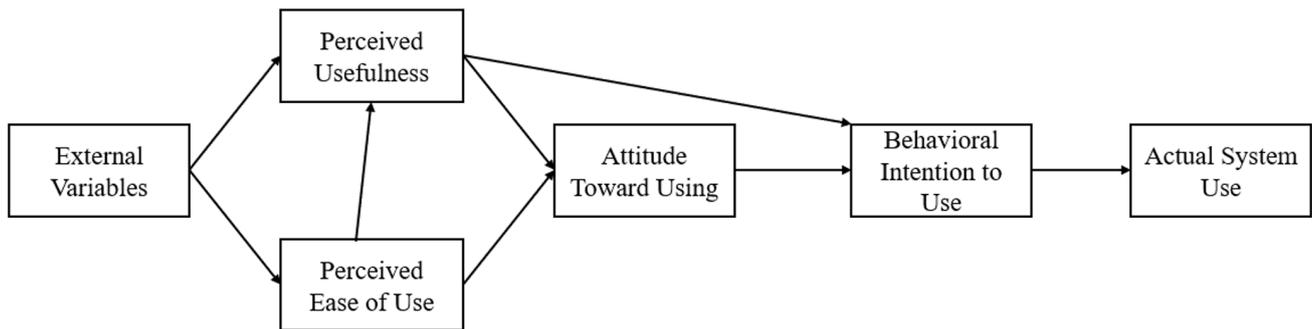


Figure 4: Elements of the Technology Acceptance Model, adapted from Davis et al. (1989, p. 985)

2.5. Hypotheses Development

Generative AI like ChatGPT can create content and perform tasks autonomously. It is a significant leap from traditional technologies due to its advanced capabilities and potential for widespread application in various domains (McKinsey, 2024). As generative AI continues to gain popularity, understanding the role of leadership style in its acceptance becomes increasingly critical. Previous studies have shown that transformational leadership as an external variable in the TAM positively influences technology acceptance (Elkhani et al., 2014; Neufeld et al., 2007) while transactional leadership does not significantly affect technology acceptance (Schepers et al., 2005). Unlike traditional technologies that often require extensive user input and learning, generative AI can enhance perceived ease of use by providing intuitive interfaces (Davenport & Ronanki, 2018). However, despite these benefits, there is still skepticism and hesitation about the use of such technologies. Studies show that concerns about the reliability, security, and ethical implications of AI applications can hold back many potential users (Kamila & Jasrotia, 2023; Smith & Anderson, 2014).

The distinctive nature of generative AI thus requires leaders who can effectively manage and leverage its potential for innovation. Transformational leadership has been proven to positively influence perceived usefulness while transactional leadership did not show any significant effects (Schepers et al., 2005). Research by Bunjak et al. (2022) indicates that transformational leadership can persuade followers to lead themselves towards achieving digital change, suggesting that this leadership style is particularly suited to environments requiring adaptability and innovation. Similarly, Aziz et al. (2020) and Molino et al. (2021) found that transformational leadership significantly influences the acceptance of new technologies.

Therefore I hypothesize:

H1: The acceptance of generative AI in the workplace is greater under transformational leadership than under transactional leadership.

Transformational leadership is proven to enhance organizational commitment (see Table 1). Organizational commitment, particularly affective commitment, is critical to creating positive attitudes toward innovation (Jafri, 2010; Herri

& Brunetto, 2013). Employees with high commitment are therefore more likely to accept the introduction and use of AI in the workplace. Hence, organizational commitment acts as a mediator between leadership style and AI acceptance. Transformational leadership increases commitment and promotes AI acceptance, while transactional leadership does not increase organizational commitment, which in turn does not increase AI acceptance.

Thus, I hypothesize:

H2: Organizational commitment mediates the relationship between leadership style and acceptance of AI in the workplace.

The threat-rigidity theory suggests that perceived threats lead to rigidity in thinking and behavior. Employees facing the threat of job loss due to AI may experience a narrowing of attention and a constriction in control, leading them to resist changes and innovations. This rigidity can significantly hinder the adoption of new technologies and processes in the workplace. Transformational leadership is characterized by providing a clear vision and direction (Bass & Avolio, 1990), building a trustful and supportive environment (Avolio et al., 2004), encouraging innovative thinking (Gumusluoglu & Ilsev, 2009; Khalili, 2016; Matzler et al., 2008). AI can bring many opportunities to employees but can also spark the fear of job losses due to automation (McClure, 2018; Ore & Sposato, 2022).

According to COR theory (Hobfoll, 1989), individuals are motivated to protect and acquire resources, such as job security. The introduction of AI can be perceived as a threat to this valuable resource, leading to resistance and rigidity in behavior as employees strive to protect what they have. By framing AI as a resource gain (e.g., increased efficiency, new career opportunities) transformational leaders can reduce the perception of threat and promote flexibility and openness among employees which then leads to a reduced fear of job loss. Employees with a lower fear of job loss are more likely to accept and engage in innovative behaviors (J. Chen et al., 2016). High levels of fear of job loss can negatively impact innovative work behavior. Employees facing fear of job loss may perceive innovations as a threat, leading to resistance and decreased acceptance (Van Hootegem et al., 2019). König et

al. (2010) demonstrated that the fear of job loss negatively impacted employees' willingness to accept and adopt new technologies in the workplace. Employees with higher job security and therefore lower fear of job loss are more likely to engage in innovative behaviors and accept organizational changes (Probst et al., 2007).

Consequently, AI acceptance is likely to be higher among employees who are less concerned about job loss due to transformational leadership. Therefore I hypothesize:

H3: The relationship between leadership style and acceptance of AI in the workplace is mediated by employee's fear of job loss.

In SET it is stated that relationships are formed through a series of interactions, where individuals weigh the costs and benefits of these interactions. In the context of leadership and organizational change, this theory suggests that employees are more likely to accept and support changes when they perceive a positive exchange relationship with their leaders. Transformational leadership can create a strong positive exchange relationship between leaders and followers. These LMX relationships are characterized by trust, respect, and mutual obligation, which can be fostered and strengthened through frequent interactions (Gajendran & Joshi, 2012; Kacmar et al., 2003).

As hypothesized in H1, transformational leaders can positively influence employees' willingness to accept AI technologies in the workplace. The frequency of interactions between leaders and followers is likely to amplify this effect (Kuvaas et al., 2017). A high interaction frequency can lead to more opportunities for leaders to demonstrate transformational behaviors, increased opportunities for employees to voice concerns and receive support and a stronger sense of trust and rapport between leaders and followers. As interaction frequency increases, the cumulative impact of transformational leadership behaviors becomes more pronounced, thus strengthening the exchange relationship. This increased exposure to positive leadership influences can strengthen the exchange relationship, potentially leading to greater acceptance of organizational changes such as AI implementation. Therefore, I hypothesize that:

H4: The influence of transformational leadership on the acceptance of AI in the workplace is positively moderated by the frequency of interactions with the supervisor.

Figure 5 illustrates the proposed conceptual model and highlights the interplay between (transformational) leadership, includes the moderating effect of interaction frequency, and the mediating roles of organizational commitment and the fear of job loss as determinants of AI acceptance.

3. Methodology

In this section, I will describe the methods I used, including a thorough description of the steps I took to collect and

examine the data. In order to facilitate the understanding and possible replication of my findings by other researchers, I aim to promote transparency and replicability.

3.1. Research Design

Quantitative research allows the results to be generalized to a larger population and can therefore be used to make statements about the likelihood that observed effects are representative of the overall population (Creswell, 2009). I conducted a 2 x 2 randomized experiment through an online questionnaire which allowed the manipulation of stimulating variables and observe the difference in the results. The research design was chosen because it was possible to apply previous research methodologies to the emerging field of generative AI, addressing the existing research gap. The 2 x 2 design allows a researcher to distinctively examine the different leadership styles as well as the predicted moderating effect of interaction frequency. The survey was conducted online on the Unipark/Tivian platform, had a completion rate of 98% (due to screeners and the use of the online survey platform prolific) and an average completion time of 4m 53s.

The survey started with a consent form followed by a screener question as suggested by Sheatsley (1983). Participants were asked if they had ever heard of ChatGPT. Answering "no" led to the end of the survey, as knowledge of ChatGPT is essential for continuance. If the participants answered "yes", the survey continued with an explanation of a fictional company that the participants were asked to imagine working for. A fictional company called Nexus was chosen instead of a real company to avoid bias. The description of the company did not specify the industry for the same reason.

Next, participants were randomly assigned one of two scenarios. Each scenario described the manager introducing a company-own AI tool like ChatGPT, called NexusGPT, via email. The tonality of the email differed between the scenarios. Half of the participants saw an email with elements of a transformational leadership style (A), while the other half received an email with a more transactional leadership style (B). The participants then had to answer questions about their organizational commitment to the company described. In the next part, participants were again randomly presented with one of two scenarios, completing the 2 x 2 design. The scenarios consisted of participants either communicating and interacting with their manager daily (1) or, in the second scenario, rarely (once or twice a month) (2), see Figure 6.

In the following section, participants were asked to answer questions about their perceived job (in)security. The participants were then asked to answer questions regarding their perceived ease of use and their perceived usefulness of NexusGPT to measure the technology acceptance.

To test whether participants recognized the frequency scenarios presented, they were asked to identify the scenario they had seen as well as rate the realism. Finally, demographic information (age, education, gender) was collected from the participants. After completing the survey, participants did not receive any information about the other email

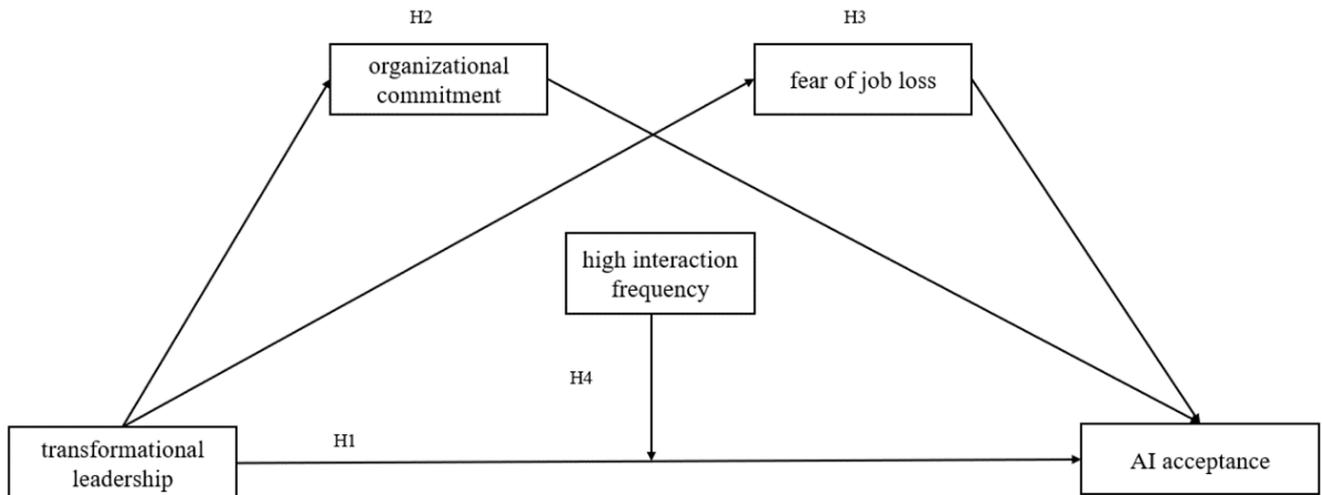


Figure 5: Conceptual model of the hypotheses

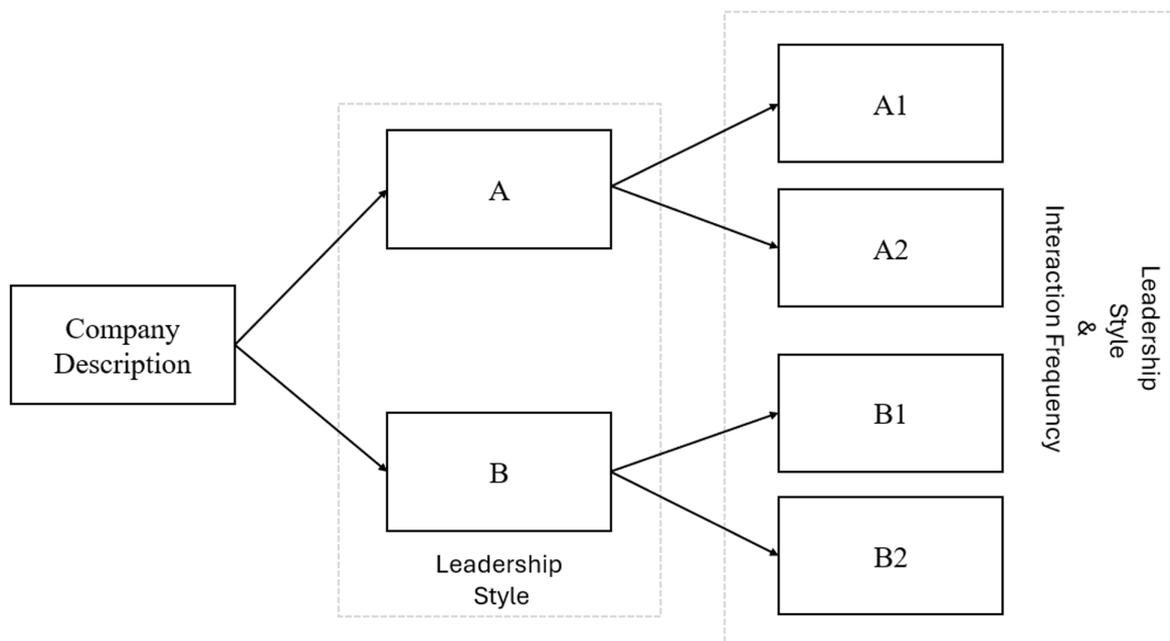


Figure 6: Visualization of the random paths

scenario or the research question. The complete questionnaire with every question can be found in Appendix J.

3.2. Measurements

Scenarios

Each participant was asked to imagine them opening their mailbox and seeing the following text as a mail from their manager. Participants who got the scenario with *transformational leadership* assigned were presented with a manager who framed the introduction of the AI tool NexusGPT as “an exciting opportunity that has the potential to enhance our creativity and efficiency”. The text used phrases like “sup-

port your work and spark innovation”, “sharing experiences” and “Let’s embrace this opportunity to learn and evolve together”. Feedback between the employees and between employees was encouraged. The text mentioned NexusGPT as a tool to drive the company’s mission forward. Participants who got the scenario with transactional leadership read a mail with the focus on the measurable value of NexusGPT. The manager mentioned an “estimated 15% increase in task completion speed” and a “potential 30% reduction in time spent on routine queries”. Employees who demonstrate improved performance using NexusGPT were offered the opportunity to receive a reward or promotion in the future. The text focused on performance and rewards, emphasizing the transactional leadership.

Organizational commitment

The questions were adapted from Mowday et al. (1979), but reduced from 15 questions to 6 questions in order to keep the survey as short as possible. Questions that could be irritating or confusing due to the fictional nature of the scenario were removed. This approach ensures clarity and relevance by maintaining the integrity of the original scale while adapting it to the specific context of the study. Questions like “*I am willing to put in a great deal of effort beyond that normally expected in order to help Nexus be successful*” and “*I really care about the fate of this organization*” were kept. The questions have been changed to replace the words “this organization” with the name “Nexus”. All items were recorded on a five-point Likert scale (1 = completely disagree, 5 = completely agree) see Appendix J.

Job insecurity

The questions used were taken from Vander Elst et al. (2014), who validated their questionnaire in five European countries. They included statements such as “*I feel insecure about the future of my job.*” or “*I am sure I can keep my job.*” and were not modified. All items were recorded on a five-point likert scale (1 = completely disagree, 5 = completely agree), see Appendix J.

Technology acceptance

The questions were taken from Davis et al. (1989) and slightly modified to account for the in the scenario mentioned “NexusGPT”. They included statements such as “*Using NexusGPT can make it easier to do my job.*” and “*Using NexusGPT can enhance my effectiveness on the job.*” The original seven-point likert scale was changed into a five-point likert scale (1 = completely disagree, 5 = completely agree) for consistency. As a last question, an attention test was added to this scale, so that participants were asked to select the corresponding item according to the prompt “*I am attentive and click on completely agree*”.

Demographics

Demographic questions were included in the questionnaire to assess whether attitudes were influenced by demographic variables. Participants were asked their age (in years), gender (m/f/d), and highest level of education at the end of the survey. These questions were placed at the end to allow participants to build trust, as demographic questions can be considered very personal by some individuals (Sheatsley, 1983).

Control variables

For control variables, two questions based on previous research on technology acceptance were asked to measure the

participants’ general openness to innovations and new technologies (Karahanna et al., 2006) and their prior experience (Venkatesh et al., 2003) with ChatGPT. The first question, “How open are you to new ideas and innovations in your life?” was rated on a scale of 1 to 5, where 1 is “not at all”, 3 is “neutral”, and 5 is “very open”. The second question, “How often do you use ChatGPT in your life?” was also rated on a Likert scale of 1 to 5 (1 = never, 3 = sometimes and 5 = very often).

Manipulation check and realism check

The manipulation was tested with true/false questions. The manipulation with the scenario was considered successful since with 83 of 104 (79.8%) participants who received the transformational scenario a large majority felt that their leader was inspiring and fostered shared growth. The participants were asked to remember if they were shown a scenario with frequent or rare interactions. The realism (Lienemann, 2021) of the scenario was rated on a Likert scale of one to five, with one being “completely unrealistic” and five being “completely realistic” ($M = 3.986$, $SD = 0.863$). For the entire study, all questions were coded as mandatory and could not be skipped. The composition of the variables from the items is listed in Appendix H.

3.3. Pretest

A pretest was conducted before the survey was released which is mandatory according to Porst (2014) The purpose of the pretest was to identify difficulties and obstacles before the actual experiment. The participants in the pretest were 13 students between the ages of 21 and 28. This ensured that each of the four randomized paths (see Figure 6) in the experiment was completed at least two times. The students were familiar with ChatGPT and have worked full-time (in internships) before which generally qualified them for the final survey. This is crucial because the population for the pretest must always be drawn from the same target group or population that has been defined for the final survey (Jacob et al., 2012). Additionally, three other volunteering participants belonging to the same demographic group were observed while filling out the survey and asked about their thoughts throughout the process. Participant engagement was monitored as well to identify any sections where respondents might lose interest. The role that participants were asked to imagine themselves in was afterwards adjusted slightly based on feedback to make it feel more realistic. In general, the students’ perception of the questionnaire was that it was well structured and that they had no major problems understanding it.

3.4. Sampling

The data was collected through the online platform Prolific. Platforms such as Prolific or Amazon’s Mechanical Turk (MTurk) have emerged as a widely used source of participants in the social sciences (Paolacci et al., 2010). Prolific

allowed for pre-screening, meaning that only potential participants who worked full-time and had high acceptance rates from other studies were able to participate. A total of 241 respondents from the U.S. participated in the study for monetary compensation in July 2024. I used screeners like approval ratings from previous participations and tenure to increase data quality (Arndt et al., 2022). 16 Participants who failed the two attention checks were also excluded, as there was no guarantee that they had read and understood the previous questions and descriptions. Two participants (first percentile) sped through the survey and were removed as speeders (Danatzis & Möller-Herm, 2023). Two participants who did not know ChatGPT were screened out as they could not provide appropriate estimates of perceived ease of use and perceived usefulness. Only one participant selected "other" as their gender. They were removed to allow for comparisons between males and females. The data cleaning resulted in a final sample of 220 respondents. Participants ranged in age from 18 to 67 ($M = 34.61$, $SD = 9.579$). 113 (51.36%) were male and 107 (48.64%) were female. Around half of the participants had a bachelor's degree (51.82%) while 17.27% had a master's degree and 22.27% a high school diploma / A-levels as their highest education. The rest (8.64%) reported to have either no degree, an intermediate school leaving certificate, an PhD or other forms of education.

3.5. Reliability and Validity

The variables were firstly tested for consistency reliability using Cronbach alphas and composite reliability, both indicators are listed in Table 2. Reliability is crucial for ensuring the validity of the measurement model before proceeding to evaluate the structural model (Fornell & Larcker, 1981). A commonly used rule of thumb indicates that Chronbach's alpha should be above .70 to show good internal consistency (Nunnally, 1978). For all constructs the Chronbach's alpha values were good (ranging from 0.84 to 0.92 and therefore exceeding 0.7) The analysis showed that Cronbach's alpha does not increase significantly when deleting a specific item. Composite Reliability should be above 0.6 (Bagozzi & Yi, 1988) which was also met (see Table 3). Therefore, all measured variables could be used.

4. Results

In this section, I show descriptive analysis, present the assumption tests and the corresponding results of my analysis. The analysis was conducted using STATA (version 18.0). I employed ANOVA and Structural Equation Modeling (SEM) to test the hypotheses. The STATA code used to generate the results is provided in Appendix I.

4.1. Descriptive analysis

Effects were assessed based on the correlation coefficient (r) as follows: weak correlation for $0.1 < r < 0.3$, moderate correlation for $0.3 < r < 0.5$, and strong correlation for $r >$

0.5 (Cohen, 1988). Age does not show a significant correlation with education level ($r = -0.0045$). This analysis excludes the responses of 5 participants who selected "other" because this response category cannot be logically ordered in the same ordinal ascending order as bachelor's, master's, and Ph.D. degrees. An analysis of the relationship between age and organizational commitment reveals no correlation ($r = 0.076$). Age and the fear of job loss show no relationship ($r = -0.0183$). Age and technology acceptance show no correlation as well ($r = -0.0764$). There is a weak negative relationship between age and perceived ease of use ($r = -0.1460$), indicating that the younger the participants are the easier they estimate the use of NexusGPT.

The average age of the male and female participants is similar ($M_{male} = 35.12$, $M_{female} = 34.07$). The most common response among participants (37.73%) is to use ChatGPT sometimes while 5.91% use ChatGPT never and 17.27% use ChatGPT very often. 17.27% indicate a rare use and 21.82% a use frequency between "sometimes" and "very often". Age and the use of ChatGPT do not correlate ($r = -0.08$). The data shows no correlation ($r = 0.0469$) between education (excluding individuals who selected 'other' for statistical reasons, see explanation above) and the use of ChatGPT. Almost half of the participants (47.73%) selected 5 (out of 5) as an answer to the question "How open are you to embracing new ideas and innovations in your life?" with 5 being "very open" and 3 being "neutral" and 1 being "not at all" ($M = 4.395$, $SD = 0.6432$). The correlation between age and openness to innovation is weak ($r = -0.1025$). The control variables openness to innovations have a moderate ($r = 0.402$) and prior experience with ChatGPT a strong positive correlation ($r = 0.5481$) with technology acceptance but do not correlate with leadership style. All means, standard deviations and correlations are shown in Table 3.

4.2. Inferential analysis

To test hypothesis 1 with a t-test comparing two means, several requirements must be met (Boneau, 1960; Pituch & Stevens, 2016). Each group should have at least 50 participants, which is satisfied with 104 in Group A (transformational) and 116 in Group B (transformational) ($n_{max}/n_{min} < 1.5$; $116/104 = 1.115$). The normal distribution of the dependent variable (technology acceptance) is assumed since $n > 30$ for both groups. The variances within each group should not be systematically different. The Levene's test for the dependent variable is not significant (technology acceptance: $p > 0.118$) which indicates homogeneity of variance.

To account for control variables (openness to innovations and prior experience) an ANCOVA was conducted to test H1. The overall model is statistically significant, $F(3, 216) = 37.37$, $p < 0.001$ and explains 34.17% of the variance ($R^2 = 0.3417$). Leadership style is not significant ($p = 0.664$), indicating no significant difference in technology acceptance between the two leadership styles. Therefore, H1 is not supported. However, the control variables showed significant effects: prior experience with ChatGPT ($p < 0.001$) and open-

Table 2: Chronbach's Alpha and Composite Reliability

Variable	Reference	Scale item (selection)	Cronbach's Alpha	Composite Reliability
organizational commitment	Mowday et al. (1979)	<ul style="list-style-type: none"> • I feel very little loyalty to Nexus. • I am proud to tell others that I am part of Nexus. • I really care about the fate of Nexus. 	.8436	.9352
fear of job loss	Vander Elst et al. (2014)	<ul style="list-style-type: none"> • Chances are, I will soon lose my job. • I feel insecure about the future of my job. • I think I might lose my job in the near future. 	.8992	.9419
technology acceptance PEOU + PU	Davis et al. (1989)	<ul style="list-style-type: none"> • Learning to operate Nexus-GPT is easy for me • It is easy for me to become skillful at using NexusGPT • Using NexusGPT can make it easier to do my job. 	.9290	.9762
perceived ease of use	Davis et al. (1989)	<ul style="list-style-type: none"> • Learning to operate Nexus-GPT is easy for me • It is easy for me to become skillful at using NexusGPT 	.9018	.9677
perceived usefulness	Davis et al. (1989)	<ul style="list-style-type: none"> • Using NexusGPT can make it easier to do my job. 	.9140	.9672

ness to innovation ($p = 0.0003$) both significantly positively influence technology acceptance.

Before conducting a mediation analysis, requirements like a direct effect, normality, linearity, and homoscedasticity must be met. A visualization of the normal distribution of technology acceptance and organizational commitment, along with Q-Q Plots (Appendix C, E, F, G), indicates a normal distribution with minor deviations at the upper end. Linearity is demonstrated in Figure 9 and 10 (see Appendix A and B). An analysis of fear of job loss, as shown in the Q-Q plot (see Appendix D) reveals skewness at the lower end and heavier-than-normal tails, indicating deviations from normality. However, Knief and Forstmeier (2021) argue that, contrary to intuition but supported by substantial literature, violations of the normality of residuals assumption are rarely problematic for hypothesis testing and parameter estimation. Commonly recommended solutions may pose greater risks than the issue itself. For this reason, as well as the sample size being rather large, I proceeded with the mediation analysis.

Hypothesis 2 proposed that organizational commitment mediates the relationship between leadership style and acceptance of AI in the workplace. A mediation analyses requires a direct effect between the dependent variable leadership style and the independent variable technology acceptance (Baron & Kenny, 1986). Since the effect is not signifi-

cant a mediation would not have been performed. However, recent research suggests that significant indirect effects can occur even without a significant direct effect (Hayes, 2009; MacKinnon et al., 2002; Shrout & Bolger, 2002). Therefore, despite the lack of a significant direct effect, I proceeded with the mediation analysis to explore potential indirect effects.

I used structural equation modeling (SEM) to examine the coefficients (Figure 8). The indirect effect of leadership style on organizational commitment is not significant ($b = 0.0434$, $p = 0.664$), while organizational commitment has a significant positive effect on technology acceptance ($b = 0.2611$, $p < 0.001$). Openness to innovation ($p < 0.01$) and prior experience with ChatGPT ($p < 0.001$) both have a significant positive influence on technology acceptance. The results do not support the hypothesis that organizational commitment mediates the relationship between leadership style and technology acceptance.

Hypothesis 3 which states that the relationship between leadership style and acceptance of AI in the workplace is mediated by employee's fear of job loss was also tested using a mediation analysis. Despite no significant direct effect, a mediation analysis was conducted for the same reasoning as discussed before. SEM analysis (as shown in Figure 7) confirms that leadership style significantly influences fear of job loss ($b = 0.3507$, $p < 0.01$). If leadership style switches from 0 (transformational) to 1 (transactional), fear of job loss in-

Table 3: Means, standard deviations, and correlations of variables

Variables	mean	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) technology acceptance	3.993	.614	1.000										
(2) perceived ease of use	4.03	.660	0.810***	1.000									
(3) perceived usefulness	3.974	.681	0.959***	0.610***	1.000								
(4) organizational commitment	3.52	.744	0.480***	0.353***	0.477***	1.000							
(5) Fear of job loss	2.399	.919	-0.363***	-0.315***	-0.338***	-0.286***	1.000						
(6) realism	3.986	.863	0.308***	0.208**	0.316***	0.341***	-0.172*	1.000					
(7) age	34.609	9.579	-0.076	-0.146*	-0.032	0.076	-0.018	0.135*	1.000				
(8) gender	1.486	.501	-0.058	-0.083	-0.038	-0.036	0.130	-0.069	-0.055	1.000			
(9) education	4.05	.937	0.005	0.017	-0.001	-0.123	-0.052	-0.152	0.068	-0.013	1.000		
(10) openness to innovation	4.395	.643	0.402***	0.417***	0.341***	0.284***	-0.272***	0.141*	-0.105	-0.132	0.179**	1.000	
(11) prior ChatGPT use	3.273	1.118	0.548***	0.552***	0.473***	0.304***	-0.273***	0.118	-0.080	-0.181**	-0.004	0.396**	1.000

* p < 0.05, ** p < 0.01, *** p < 0.001, n=220

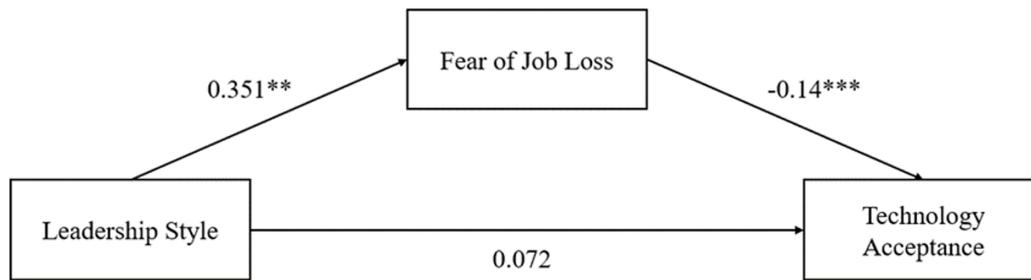


Figure 7: H3 with path coefficients

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

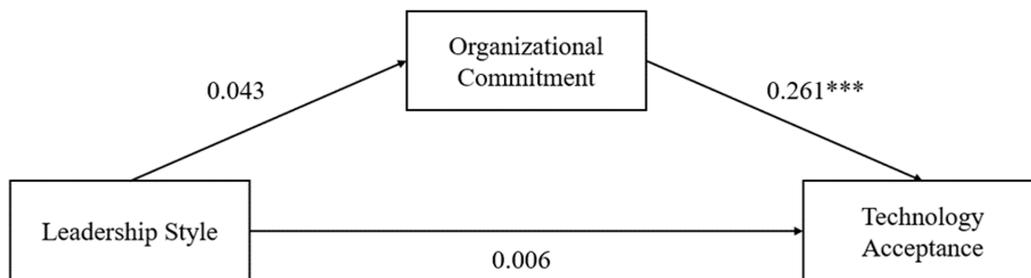


Figure 8: H2 with path coefficients

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

creases approximately by 0.35. This demonstrates that transformational leadership leads to significantly lower fear of job loss ($M_{transform} = 2.214$, $SD = 0.083$; $M_{transakt} = 2.565$, $SD = 0.088$) while fear of job loss is higher by transactional leadership. Fear of job loss has a significant negative effect on technology acceptance (coefficient = -0.14 , $p < 0.001$). If fear of job loss increases by 1 unit, technology acceptance is reduced by approximately 0.14. Openness to innovation ($p < 0.01$) and prior experience with ChatGPT ($p < 0.001$) significantly influence technology acceptance. Since the indirect effect is significant, H3 is confirmed.

A two factorial ANOVA (each group > 50) was conducted to examine the interaction effect of leadership style and interaction frequency on technology acceptance. The overall model is not statistically significant ($F(3, 216) = 0.93$, $p = 0.4278$), indicating that the predictors do not explain a significant portion of the variance in technology acceptance. The interaction between leadership style and interaction frequency (leadership style * interaction frequency) is not significant ($F(1, 216) = 1.93$, $p = 0.1664$) with an R^2 of 0.0127 explaining only 1.27% of the variance. This leads to the rejection of H4. A post-hoc pairwise analysis was not conducted since the interaction effect is not significant.

To assess the robustness of my findings, a model specification test was carried out to check for the possibility of omitted variables or model misspecification. As outlined by Hosmer et al. (2013) the goodness of link test was used. The results indicate a significant outcome. for the linear predicted value, while the squared predicted value is insignificant. These find-

ings indicate that the model is free from omitted variables and is correctly specified, further supporting its validity.

5. Discussion

In this section I discuss my empirical findings and compare them to existing literature. I resort to studies I referenced in chapter 2 (Theory and Hypotheses) and contextualize them with other research when the hypotheses are not confirmed. Additionally, I discuss the theoretical and practical contributions of this thesis, as well as its limitations.

5.1. General Discussion

The purpose of this thesis was to investigate whether leadership styles differ in their influence on the acceptance of AI in the workplace. Furthermore, the mediating effect of fear of job loss and organizational commitment as well as the moderating effect of interaction frequency were examined.

Based on the previous theoretical contributions and empirical studies, I deduced Hypothesis 1 which proposed that the acceptance of generative AI in the workplace is greater under transformational leadership than under transactional leadership. Conversely, the results did not support H1. This finding is surprising given the extensive literature suggesting that transformational leadership generally promotes higher levels of innovation and change acceptance (Bunjak et al., 2022; Elkhani et al., 2014; Neufeld et al., 2007). A possible explanation is that the inherent novelty and complexity of generative AI may require more than visionary leadership.

Employees might also demand specific technical support and reassurance about the technology's benefits and risks, which are not captured solely by leadership style. Employees may also perceive these technologies as more disruptive than traditional tools, which could diminish the hypothesized positive effect of transformational leadership.

Another explanation is that prior experience with ChatGPT facilitates acceptance in the workplace. The results showed that individuals who regularly use ChatGPT tend to have higher acceptance rates. Familiarity with the introduced technology (generative AI) plays a crucial role in acceptance. This is consistent with the findings on TAM, which indicates that user experience and familiarity with technology positively influence its acceptance (Marangunić & Granić, 2015). Openness to innovation, as a control variable, helped to explain the acceptance of generative AI. Research indicates that personal innovativeness, closely related to openness to innovation, significantly moderates the effects of key factors on the intention to use generative AI technologies. For example, Foroughi et al. (2023) found that personal innovativeness reduces the influence of performance expectancy, effort expectancy, social influence, and facilitating conditions on the intention to use ChatGPT for educational purposes. This suggests that individuals open to innovation are more inclined to experiment with new technologies regardless of other influencing factors (Foroughi et al., 2023).

Hypothesis 2 proposed that organizational commitment would mediate the relationship between leadership style and AI acceptance. The results showed that while transformational leadership did not have a significant effect on organizational commitment, organizational commitment itself had a significant effect on technology acceptance. This finding suggests that organizational commitment is a critical factor in AI acceptance, independent of the direct influence of leadership style. This result can be interpreted in several ways. First, it highlights the importance of organizational commitment as a stand-alone factor in the successful adoption of AI technologies as supported by previous studies (Vella et al., 2013). Employees who are emotionally attached to their organization and identify with its goals are more likely to support and accept new technologies such as AI, regardless of the influence of their leaders. Second, other factors like organizational culture or organizational support could moderate the influence leadership has on organizational commitment. Huey Yiing and Zaman Bin Ahmad (2009) found that innovative and supportive cultures positively influence the relationship between leadership behavior and organizational commitment. A study from Eisenberger et al. (2002) showed that perceived organizational support enhances organizational commitment and influences employees' willingness to embrace new technologies.

Based on the Threat rigidity theory and COR theory Hypothesis 3 posited that fear of job loss mediates the relationship between leadership style and AI acceptance. The findings from the mediation analysis confirmed this hypothesis. Transformational leadership significantly reduced fear of job loss, which in turn increased AI acceptance. This is consis-

tent with threat rigidity theory (Staw et al., 1981) and highlights the importance of addressing employees' fears about job security in the context of AI introduction. By integrating COR theory with threat rigidity theory, this thesis offers a more nuanced understanding of the psychological mechanisms employees experience when facing potential job loss. It examines how the threat of resource loss, driven by AI-related automation and job displacement, impacts employees' responses. Transformational leaders can mitigate these fears and increase acceptance of new technologies by providing a clear vision and fostering a supportive environment. This aligns with previous research indicating that transformational leaders can mitigate employees' anxiety and resistance to change (Bass & Avolio, 1993; Rafferty & Griffin, 2004). Other factors, such as job type (Keim et al., 2014) and training (Kalleberg, 2009), may also play a role in further reducing fear of job loss. Longer job tenure can provide employees with a greater sense of stability and organizational commitment, while comprehensive training programs can equip them with the skills they need to feel confident and prepared for new technologies (Venkatesh et al., 2012).

Hypothesis 4 proposed that the frequency of interactions with the supervisor would moderate the influence of transformational leadership on AI acceptance. However, the data did not support this hypothesis. The frequency of interactions with the supervisor did not significantly enhance the assumed positive effects of transformational leadership on AI acceptance. This finding suggests that while frequent interactions may generally benefit leader-follower relationships (Gajendran & Joshi, 2012), they do not necessarily amplify the specific impact of transformational leadership on technology acceptance. One possible explanation is that other factors, such as the length and depth of interactions, play a much more important role than interaction frequency. Therefore, while simply increasing the number of interactions between a leader and their followers might not be sufficient to enhance AI acceptance, focusing on the quality, content, and context of these interactions could potentially yield better results. A second explanation could be that the familiarity gained through frequent interactions does not necessarily translate into increased influence over complex attitudinal changes such as AI acceptance. While frequent interactions may build general rapport and trust, they may not provide the detailed, high-quality exchanges potentially needed to address AI-related concerns.

5.2. Theoretical Contributions

This thesis contributes to the existing literature on leadership and technology acceptance in the context of generative AI in several ways. First, it extends the research regarding threat rigidity theory (Staw et al., 1981) by demonstrating how transformational leadership can reduce employees' fears of job loss and thus facilitate the acceptance of AI technologies. Previous studies have focused on the effect of leadership styles without considering the specific threats posed by advanced technologies such as AI. By incorporating these elements, these findings add a new dimension to

the understanding of threat rigidity and leadership. Second, this thesis empirically demonstrates the importance of considering multiple factors, beyond leadership style, in understanding AI acceptance. Specifically, it highlights how factors such as prior experience with AI technologies, organizational commitment, and openness to innovation play important roles in fostering AI acceptance. Third, building on the transformational leadership theory (Bass & Avolio, 1993), this study clarifies the theory's role in enhancing technology acceptance. While transformational leadership is well-documented for its positive effects on organizational commitment and innovation (Gumusluoglu & Ilsev, 2009; Khalili, 2016), these findings specifically highlight its effectiveness in reducing fear of job loss and consequently improving AI acceptance. This contributes to a better understanding of how transformational leadership can be leveraged to manage technological disruptions in the workplace. Finally, although the findings do not support the proposed relationship between leadership style and organizational commitment, they do identify organizational commitment as a significant predictor of technology acceptance. This enriches the current research of factors influencing the acceptance of AI (Kelly et al., 2023).

5.3. Practical Implications

The results of this thesis provide valuable insights for managers planning to implement generative AI in the workplace. Overall, leadership plays a significant role in determining the final AI acceptance. If leadership fails to address the concerns and fears associated with AI, it may lead to significant resistance and low technology acceptance among employees. This can potentially jeopardize the implementation efforts. Managers should consider three key aspects when introducing AI. First, they should aim to adopt a transformational leadership style. Transformational leaders can effectively reduce employees' fear of job loss by providing a clear vision for the future, inspiring confidence, and offering continuous support. The results of this thesis show, that this reduced fear of job loss leads to a higher AI acceptance. Second, managers should look for ways to improve organizational commitment since it has been proven to increase AI acceptance. One way to do this is to develop a better corporate culture (Abdul Rashid et al., 2003; Clugston et al., 2000). Another way is to provide opportunities for professional development and career advancement, which can foster a sense of loyalty and attachment to the organization (Meyer & Allen, 1991). Third, Managers should also incorporate the fact that openness to innovation has a significant effect on technology acceptance. Before introducing AI in the workplace, they should foster a culture that encourages curiosity and continuous learning. Managers can achieve this through training programs, open communication and involving employees in the implementation.

5.4. Limitations and Future Research

This thesis has some limitations that might be overcome by future research to strengthen the design of this study.

First, the research design relied on self-reported data and a fictional company. Further research could apply the same theoretical model using naturalistic methods such as field experiments in a real company. This could improve the homogeneity of the sample and result in less noisy data. It could also enhance the participants ability to give profound answers regarding their organizational commitment. Second, the experiment used an e-mail as a stimulus to introduce AI to the workplace. The stimulus as an e-mail could have potentially undermined the inspirational effect transformational leaders have. Although corporate changes are often communicated digitally (Ruck & Welch, 2012), it does not necessarily have to be via e-mail. Further research could examine whether the channel (e-mail, online video, voice memo or face-to-face announcement) of AI introduction might moderate the relationship between leadership style and technology acceptance.

Third, the participants in the experiment were all U.S. Americans and not differentiated by industry or company size. Prior personal experiences in a large tech company might significantly differ from experiences gained in the fashion industry, consulting or the public sector. The industry might influence the perceived job security, with sectors like creative industries being more prone to concerns about job displacement due to the rise of generative AI. Further research could explore the role of industry or culture in facilitating technology acceptance, organizational commitment, and reduced fear of job loss. Fourth, since only quantitative methods were used in this thesis, detailed insight into different reasons for individuals' perceptions of fear of job loss cannot be examined. Further research could implement a mixed method approach where quantitative data is used to identify patterns and trends. Qualitative methods, such as interviews or focus groups, could then be employed to explore the underlying reasons, personal experiences, and contextual factors that contribute to individuals' perceptions of fear of job loss. It is possible that our current understanding is insufficient to conduct a deductive study like this one. It may be necessary to first revisit the topic with an inductive approach, such as grounded theory (Glaser & Strauss, 1967), to more precisely identify the potential influencing factors.

Finally, I only used transactional and transformational leadership as stimuli. Although these leadership styles are popular, there are several other leadership styles, such as situational leadership (Hersey & Blanchard, 1969), servant leadership (Greenleaf, 1977) or democratic leadership (Lewin et al., 1939). Further research could broaden the scope of leadership styles to investigate whether other leadership styles might facilitate AI acceptance better.

6. Conclusion

The aim of this thesis is to shed light on the influence leadership style has on the acceptance of AI in the workplace. Contrary to expectations, transformational leadership does not significantly enhance AI acceptance compared to transactional leadership. This suggests that the complex and

novel nature of generative AI might demand more than visionary leadership. Practical considerations such as the openness to innovations appear crucial. Regular use of ChatGPT is strongly correlated with higher AI acceptance which indicates that user experience plays a vital role in shaping attitudes toward new technologies. Furthermore, my analysis reveals that organizational commitment significantly influences AI acceptance, independent of leadership style. This highlights the importance of fostering a supportive organizational culture that encourages commitment, which then facilitates the integration of advanced technologies like generative AI. Additionally, the reduction of job loss fears through transformational leadership is confirmed as a significant factor in enhancing AI acceptance. This finding underscores the necessity for leaders to address employee anxieties to drive successful technological adoption. While the study contributes to the understanding of leadership's role in technology acceptance, it also opens areas for future research. Specifically, exploring other leadership styles and their impact on AI acceptance could provide a more comprehensive understanding. Moreover, investigating the role of industry context and the influence of different communication channels in AI implementation may offer further insights. These directions could deepen our understanding of the interplay between leadership, organizational dynamics, and the acceptance of generative AI.

References

- Abdul Rashid, Z., Sambasivan, M., & Johari, J. (2003). The influence of corporate culture and organisational commitment on performance. *Journal of Management Development*, 22(8), 708–728. <https://doi.org/10.1108/02621710310487873>
- Abdullah, F., Ward, R., & Ahmed, E. (2016). Investigating the influence of the most commonly used external variables of TAM on students' Perceived Ease of Use (PEOU) and Perceived Usefulness (PU) of e-portfolios. *Computers in Human Behavior*, 63, 75–90. <https://doi.org/10.1016/j.chb.2016.05.014>
- Aboelimged, M., & Gebba, T. R. (2013). Mobile Banking Adoption: An Examination of Technology Acceptance Model and Theory of Planned Behavior. *International Journal of Business Research and Development*, 2(1). <https://doi.org/10.24102/ijbrd.v2i1.263>
- Ajzen, I. (1980). *Understanding Attitudes and Predicting Social Behavior*. Prentice-Hall. <https://cir.nii.ac.jp/crid/1572543024551612928>
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Arndt, A. D., Ford, J. B., Babin, B. J., & Luong, V. (2022). Collecting samples from online services: How to use screeners to improve data quality. *International Journal of Research in Marketing*, 39(1), 117–133. <https://doi.org/10.1016/j.ijresmar.2021.05.001>
- Avolio, B. J., Zhu, W., Koh, W., & Bhatia, P. (2004). Transformational leadership and organizational commitment: Mediating role of psychological empowerment and moderating role of structural distance. *Journal of Organizational Behavior*, 25(8), 951–968. <https://doi.org/10.1002/job.283>
- Aziz, F., Rami, A., Razali, F., & Mahadi, N. (2020). The Influence of Leadership Style Towards Technology Acceptance in Organization. *International Journal of Advanced Science and Technology*, 29(7).
- Bagozzi, R. (2007). The Legacy of the Technology Acceptance Model and a Proposal for a Paradigm Shift. *Journal of the Association for Information Systems*, 8(4), 244–254. <https://doi.org/10.17705/1jais.00122>
- Bagozzi, R., & Yi, Y. (1988). On the evaluation of structural equation models. *Bakari, H., Hunjra, A. I., & Niazi, G. S. K. (2017). How Does Authentic Leadership Influence Planned Organizational Change? The Role of Employees' Perceptions: Integration of Theory of Planned Behavior and Lewin's Three Step Model. Journal of Change Management. https://www.tandfonline.com/doi/abs/10.1080/14697017.2017.1299370*
- Barki, H., & Benbasat, I. (2007). Quo vadis TAM? *Journal of the Association for Information Systems*, 8(4). <https://doi.org/10.17705/1jais.00126>
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- Barrett, G. V., Thornton, C. L., & Cabe, P. A. (1968). Human Factors Evaluation of a Computer Based Information Storage and Retrieval System. *Human Factors*, 10(4), 431–436. <https://doi.org/10.1177/001872086801000414>
- Bass, B. M. (1985). Leadership and performance beyond expectations, by Bernard M. Bass. *Human Resource Management*, 25(3), 481–484. <https://doi.org/10.1002/hrm.3930250310>
- Bass, B. M. (1999). Two Decades of Research and Development in Transformational Leadership. *European Journal of Work and Organizational Psychology*, 8(1), 9–32. <https://doi.org/10.1080/135943299398410>
- Bass, B. M., & Avolio, B. J. (1990). The implications of transactional and transformational leadership for individual, team, and organizational development. *Research in Organizational Change and Development*, 4(1), 231–272.
- Bass, B. M., & Avolio, B. J. (1993). Transformational Leadership and Organizational Culture. *Public Administration Quarterly*, 17(1), 112–121.
- Blau, P. (1964). *Exchange and Power in Social Life* (2nd). Routledge. <https://doi.org/10.4324/9780203792643>
- Boneau, C. A. (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, 57(1), 49–64. <https://doi.org/10.1037/h0041412>
- Buabeng-Andoh, C. (2018). Predicting students' intention to adopt mobile learning: A combination of theory of reasoned action and technology acceptance model. *Journal of Research in Innovative Teaching & Learning*, 11(2), 178–191. <https://doi.org/10.1108/JRIT-03-2017-0004>
- Bunjak, A., Bruch, H., & Černe, M. (2022). Context is key: The joint roles of transformational and shared leadership and management innovation in predicting employee IT innovation adoption. *International Journal of Information Management*, 66, 102516. <https://doi.org/10.1016/j.ijinfomgt.2022.102516>
- Burns, J. M. (1978). *Leadership*. Open Road Media.
- Burton-Jones, A., & Hubona, G. S. (2006). The mediation of external variables in the technology acceptance model. *Information & Management*, 43(6), 706–717. <https://doi.org/10.1016/j.im.2006.03.007>
- Cacioppo, J. T., & Gardner, W. L. (1999). EMOTION. *Annual Review of Psychology*, 50, 191–214. <https://doi.org/10.1146/annurev.psych.50.1.191>
- Canziani, B., & MacSween, S. (2021). Consumer acceptance of voice-activated smart home devices for product information seeking and online ordering. *Computers in Human Behavior*, 119, 106714. <https://doi.org/10.1016/j.chb.2021.106714>
- Cascella, M., Montomoli, J., Bellini, V., & Bignami, E. (2023). Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *Journal of Medical Systems*, 47(1), 33. <https://doi.org/10.1007/s10916-023-01925-4>
- Castiblanco Jimenez, I. A., Cepeda García, L. C., Violante, M. G., Marcolin, E., & Vezzetti, E. (2021). Commonly Used External TAM Variables in e-Learning, Agriculture and Virtual Reality Applications. *Future Internet*, 13(1), Article 1. <https://doi.org/10.3390/fi13010007>
- Chen, J., Leung, W. S., & Evans, K. P. (2016). Are Employee-Friendly Workplaces Conducive to Innovation? [SSRN Scholarly Paper 2662260]. <https://doi.org/10.2139/ssrn.2662260>

- Chen, L. Y. (2004). Examining the effect of organization culture and leadership behaviors on organizational commitment, job satisfaction, and job performance at small and middle-sized firms of Taiwan. *Journal of American Academy of Business Cambridge*. <https://www.sid.ir/paper/581965/en>
- Chen, S.-C., & Li, S.-H. (2010). Consumer adoption of e-service: Integrating technology readiness with the theory of planned behavior.
- Chughtai, A. A., & Zafar, S. (2006). Antecedents and Consequences of Organizational Commitment Among Pakistani University Teachers.
- Clugston, M., Howell, J. P., & Dorfman, P. W. (2000). Does Cultural Socialization Predict Multiple Bases and Foci of Commitment? *JOURNAL OF MANAGEMENT*, 26(1).
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd). Routledge. <https://doi.org/10.4324/9780203771587>
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd). Sage Publications.
- Cropanzano, R., Anthony, E. L., Daniels, S. R., & Hall, A. V. (2017). Social Exchange Theory: A Critical Review with Theoretical Remedies. *Academy of Management Annals*, 11(1), 479–516. <https://doi.org/10.5465/annals.2015.0099>
- Cropanzano, R., & Mitchell, M. S. (2005). Social Exchange Theory: An Interdisciplinary Review. *Journal of Management*, 31(6), 874–900. <https://doi.org/10.1177/0149206305279602>
- Dale, K., & Fox, M. L. (2008). Leadership Style and Organizational Commitment: Mediating Effect of Role Stress. *Journal of Managerial Issues*, 20(1), 109–130.
- Danatzi, I., & Möller-Herm, J. (2023). Stopping the Spread: How Blame Attributions Drive Customer-to-Customer Misbehavior Contagion and What Frontline Employees Can Do to Curb It. *Journal of Service Research*, 26(3), 459–475. <https://doi.org/10.1177/10946705221150441>
- Davenport, T., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116.
- Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* [Doctoral dissertation, Massachusetts Institute of Technology]. <https://dspace.mit.edu/bitstream/handle/1721.1/15192/14927137-MIT.pdf>
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35(8), 982–1003. <https://doi.org/10.1287/mnsc.35.8.982>
- Deloitte. (2023). AI study: Over 60 per cent use Artificial Intelligence at work – almost half of all employees are worried about losing their jobs. <https://www2.deloitte.com/ch/en/pages/press-releases/articles/ai-study-almost-half-of-all-employees-are-worried-about-losing-their-jobs.html>
- Eisenbach, R., Watson, K., & Pillai, R. (1999). Transformational leadership in the context of organizational change. *Journal of Organizational Change Management*, 12(2), 80–89. <https://doi.org/10.1108/09534819910263631>
- Eisenberger, R., Stinglhamber, F., Vandenberghe, C., Sucharski, I. L., & Rhoades, L. (2002). Perceived supervisor support: Contributions to perceived organizational support and employee retention. *Journal of Applied Psychology*, 87(3), 565–573. <https://doi.org/10.1037/0021-9010.87.3.565>
- Elkhani, N., Soltani, S., & Nazir Ahmad, M. (2014). The effects of transformational leadership and ERP system self-efficacy on ERP system usage. *Journal of Enterprise Information Management*, 27(6), 759–785. <https://doi.org/10.1108/JEIM-06-2013-0031>
- Erkutlu, H. (2008). The impact of transformational leadership on organizational and leadership effectiveness: The Turkish case. *Journal of Management Development*, 27(7), 708–726. <https://doi.org/10.1108/02621710810883616>
- Faupel, S., & Stüß, S. (2019). The Effect of Transformational Leadership on Employees During Organizational Change – An Empirical Analysis. *Journal of Change Management*, 19(3), 145–166. <https://doi.org/10.1080/14697017.2018.1447006>
- Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research* (Vol. 27).
- Ford, M. T., Heinen, B. A., & Langkamer, K. L. (2007). Work and Family Satisfaction and Conflict: A Meta-Analysis of Cross-Domain Relations. *Journal of Applied Psychology*, 92(1), 57–80. <https://doi.org/10.1037/0021-9010.92.1.57>
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1). <https://doi.org/10.1177/002224378101800104>
- Foroughi, B., Senali, M. G., Iranmanesh, M., Khanfar, A., Ghobakhloo, M., Annamalai, N., & Naghmehe-Abbaspour, B. (2023). Determinants of Intention to Use ChatGPT for Educational Purposes: Findings from PLS-SEM and fsQCA. *International Journal of Human-Computer Interaction*, 0(0), 1–20. <https://doi.org/10.1080/10447318.2023.2226495>
- Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, Challenges, and AI-Human Collaboration. *Journal of Information Technology Case and Application Research*, 25(3), 277–304. <https://doi.org/10.1080/15228053.2023.2233814>
- Gajendran, R. S., & Joshi, A. (2012). Innovation in Globally Distributed Teams: The Role of LMX, Communication Frequency, and Member Influence on Team Decisions. *Journal of Applied Psychology*, 97(6), 1252–1261. <https://doi.org/10.1037/a0028958>
- Glaser, B., & Strauss, A. (1967). The Discovery of Grounded Theory.
- Gouldner, A. W. (1960). The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review*, 25(2), 161–178. <https://doi.org/10.2307/2092623>
- Graen, G., & Cashman, J. F. (1975). A Role-Making Model of Leadership in Formal Organizations: A Developmental Approach. *Leadership Frontiers*, 86–96.
- Graen, G. B., & Uhl-Bien, M. (1995). Relationship-Based Approach to Leadership: Development of Leader-Member Exchange (LMX) Theory of Leadership over 25 Years: Applying a Multi-Level Multi-Domain Perspective. *The Leadership Quarterly*, 6(2), 219–247. [https://doi.org/10.1016/1048-9843\(95\)90036-5](https://doi.org/10.1016/1048-9843(95)90036-5)
- Greenleaf, R. K. (1977). *Servant Leadership: A Journey Into the Nature of Legitimate Power and Greatness*. Paulist Press.
- Gu, Y., Diao, F., & Cao, X. (2024). The Impact of Empowering Leadership on Employees' Deviant Innovation Behavior from the Perspective of Planned Behavior. In Y. P. Tu & M. Chi (Eds.), *E-Business. New Challenges and Opportunities for Digital-Enabled Intelligent Future* (pp. 133–146). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-60260-3_12
- Gumusluoglu, L., & Ilsev, A. (2009). Transformational Leadership, Creativity, and Organizational Innovation. *Journal of Business Research*, 62(4), 461–473. <https://doi.org/10.1016/j.jbusres.2007.07.032>
- Halbesleben, J. R. B., Neveu, J.-P., Paustian-Underdahl, S. C., & Westman, M. (2014). Getting to the “COR”: Understanding the Role of Resources in Conservation of Resources Theory. *Journal of Management*, 40(5), 1334–1364. <https://doi.org/10.1177/0149206314527130>
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium. *Communication Monographs*, 76(4), 408–420. <https://doi.org/10.1080/03637750903310360>
- Herold, D. M., Fedor, D. B., Caldwell, S., & Liu, Y. (2008). The Effects of Transformational and Change Leadership on Employees' Commitment to a Change: A Multilevel Study. *Journal of Applied Psychology*, 93(2), 346–357. <https://doi.org/10.1037/0021-9010.93.2.346>
- Herscovitch, L., & Meyer, J. P. (2002). Commitment to Organizational Change: Extension of a Three-Component Model. *Journal of Applied Psychology*, 87(3), 474–487. <https://doi.org/10.1037/0021-9010.87.3.474>
- Hersey, P., & Blanchard, K. H. (1969). Management of Organizational Behavior. *Academy of Management Journal*, 12(4), 526–526. <https://doi.org/10.5465/amj.1969.19201155>
- Hobfoll, S. E. (1989). Conservation of Resources: A New Attempt at Conceptualizing Stress. *American Psychologist*, 44(3), 513–524. <https://doi.org/10.1037/0003-066X.44.3.513>

- Hobfoll, S. E. (2001). The Influence of Culture, Community, and the Nested-Self in the Stress Process: Advancing Conservation of Resources Theory. *Applied Psychology*, 50(3), 337–421. <https://doi.org/10.1111/1464-0597.00062>
- Hobfoll, S. E., Halbesleben, J., Neveu, J.-P., & Westman, M. (2018). Conservation of Resources in the Organizational Context: The Reality of Resources and Their Consequences. *Annual Review of Organizational Psychology and Organizational Behavior*, 5(1), 103–128. <https://doi.org/10.1146/annurev-orgpsych-032117-104640>
- Homans, G. C. (1958). Social Behavior as Exchange. *American Journal of Sociology*, 63(6), 597–606. <https://doi.org/10.1086/222355>
- Homans, G. C. (1961). *Social Behavior: Its Elementary Forms* (Vol. 63). Harcourt, Brace; World.
- Hosmer, D., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley & Sons.
- Huey Yiing, L., & Zaman Bin Ahmad, K. (2009). The Moderating Effects of Organizational Culture on the Relationships between Leadership Behaviour and Organizational Commitment and between Organizational Commitment and Job Satisfaction and Performance. *Leadership & Organization Development Journal*, 30(1), 53–86. <https://doi.org/10.1108/01437730910927106>
- Ismael, A., Mohamed, H. A. B., Sulaiman, A. Z., Mohamad, M., & Yusuf, M. H. (2011). An Empirical Study of the Relationship between Transformational Leadership, Empowerment and Organizational Commitment. *Business and Economics Research Journal*, 2(1), 89.
- Ivanov, S., Soliman, M., Tuomi, A., Alkathiri, N. A., & Al-Alawi, A. N. (2024). Drivers of Generative AI Adoption in Higher Education through the Lens of the Theory of Planned Behaviour. *Technology in Society*, 77, 102521. <https://doi.org/10.1016/j.techsoc.2024.102521>
- Iverson, R. D. (1996). Employee acceptance of organizational change: the role of organizational commitment. *International Journal of Human Resource Management*, 7(1), 122–149.
- Jacob, R., Heinz, A., Décioux, J. P., & Eirmbter, W. H. (2012). *Umfrage: Einführung in die Methoden der Umfrageforschung*. Oldenbourg Wissenschaftsverlag. <https://doi.org/10.1524/9783486710090>
- Jafri, M. H. (2010). Organizational Commitment and Employee's Innovative Behavior: A Study in Retail Sector. *Journal of Management and Research*. <https://www.semanticscholar.org/paper/Organizational-Commitment-and-Employee-s-Innovative-Jafri/4fd06cc153678de5e7656b70f18c3e7f0174af70>
- Judge, T. A., & Bono, J. E. (2000). Five-Factor Model of Personality and Transformational Leadership. *Journal of Applied Psychology*, 85(5), 751–765. <https://doi.org/10.1037/0021-9010.85.5.751>
- Kacmar, K. M., Witt, L. A., Zivnuska, S., & Gully, S. M. (2003). The Interactive Effect of Leader-Member Exchange and Communication Frequency on Performance Ratings. *Journal of Applied Psychology*, 88(4), 764–772. <https://doi.org/10.1037/0021-9010.88.4.764>
- Kalleberg, A. L. (2009). Precarious Work, Insecure Workers: Employment Relations in Transition. *American Sociological Review*, 74(1), 1–22. <https://doi.org/10.1177/000312240907400101>
- Kamila, M. K., & Jasrotia, S. S. (2023). Ethical Issues in the Development of Artificial Intelligence: Recognizing the Risks [ahead-of-print]. *International Journal of Ethics and Systems*. <https://doi.org/10.1108/IJOES-05-2023-0107>
- Karahanna, E., Agarwal, R., & Angst, C. M. (2006). Reconceptualizing Compatibility Beliefs in Technology Acceptance Research. *MIS Quarterly*, 30(4), 781–804. <https://doi.org/10.2307/25148754>
- Keim, A. C., Landis, R. S., Pierce, C. A., & Earnest, D. R. (2014). Why Do Employees Worry About Their Jobs? A Meta-Analytic Review of Predictors of Job Insecurity. *Journal of Occupational Health Psychology*, 19(3), 269–290. <https://doi.org/10.1037/a0036743>
- Kelly, S., Kaye, S.-A., & Oviedo-Trespalacios, O. (2023). What Factors Contribute to the Acceptance of Artificial Intelligence? A Systematic Review. *Telematics and Informatics*, 77, 101925. <https://doi.org/10.1016/j.tele.2022.101925>
- Khalili, A. (2016). Linking Transformational Leadership, Creativity, Innovation, and Innovation-Supportive Climate. *Management Decision*, 54(9), 2277–2293. <https://doi.org/10.1108/MD-03-2016-0196>
- Khogali, H. O., & Mekid, S. (2023). The Blended Future of Automation and AI: Examining Some Long-Term Societal and Ethical Impact Features. *Technology in Society*, 73, 102232. <https://doi.org/10.1016/j.techsoc.2023.102232>
- Knief, U., & Forstmeier, W. (2021). Violating the Normality Assumption May Be the Lesser of Two Evils. *Behavior Research Methods*, 53(6), 2576–2590. <https://doi.org/10.3758/s13428-021-01587-5>
- Koberg, C. S., Detienne, D. R., & Heppard, K. A. (2003). An Empirical Test of Environmental, Organizational, and Process Factors Affecting Incremental and Radical Innovation. *The Journal of High Technology Management Research*, 14(1), 21–45. [https://doi.org/10.1016/S1047-8310\(03\)00003-8](https://doi.org/10.1016/S1047-8310(03)00003-8)
- König, C. J., Debus, M. E., Häusler, S., Lendenmann, N., & Kleinmann, M. (2010). Examining Occupational Self-Efficacy, Work Locus of Control and Communication as Moderators of the Job Insecurity—Job Performance Relationship. *Economic and Industrial Democracy*, 31(2), 231–247. <https://doi.org/10.1177/0143831X09358629>
- KPMG. (2024). Erwartete positive und negative Effekte durch Nutzung von generativer KI in Deutschland im Jahr 2024 [Published May 12, 2024]. *Statista*. <https://de.statista.com/statistik/daten/studie/1474797/umfrage/erwartete-positive-und-negative-effekte-durch-nutzung-von-generativer-ki/>
- Kuhnert, K. W., & Lewis, P. (1987). Transactional and Transformational Leadership: A Constructive/Developmental Analysis. *Academy of Management Review*, 12(4), 648–657. <https://doi.org/10.5465/amr.1987.4306717>
- Kuvaas, B., Buch, R., & Dysvik, A. (2017). Constructive Supervisor Feedback is not Sufficient: Immediacy and Frequency is Essential. *Human Resource Management*, 56(3), 519–531. <https://doi.org/10.1002/hrm.21785>
- Lee, J., Cerreto, F. A., & Lee, J. (2010). Theory of Planned Behavior and Teachers' Decisions Regarding Use of Educational Technology. *Educational Technology & Society*, 13(1), 152–164.
- Lewin, K., Lippitt, R., & White, R. K. (1939). Patterns of Aggressive Behavior in Experimentally Created “Social Climates”. *The Journal of Social Psychology*. <https://www.tandfonline.com/doi/abs/10.1080/00224545.1939.9713366>
- Lienemann, A. J. (2021). Ergebnisse der empirischen Untersuchung. In A. J. Lienemann (Ed.), *Die Wirkung von Beziehungen im Influencer Branding: Eine Experimentelle Analyse des Brand-Fits und des Influencer-Attachments* (pp. 135–158). Springer Fachmedien. https://doi.org/10.1007/978-3-658-35467-1_11
- Limsila, K., & Ogunlana, S. O. (2008). Performance and Leadership Outcome Correlates of Leadership Styles and Subordinate Commitment. *Engineering, Construction and Architectural Management*, 15(2), 164–184. <https://doi.org/10.1108/09699980810852682>
- Lu, Y., Zhou, T., & Wang, B. (2009). Exploring Chinese Users' Acceptance of Instant Messaging using the Theory of Planned Behavior, the Technology Acceptance Model, and the Flow Theory. *Computers in Human Behavior*, 25(1), 29–39. <https://doi.org/10.1016/j.chb.2008.06.002>
- Lyons, J. B., & Schneider, T. R. (2009). The Effects of Leadership Style on Stress Outcomes. *The Leadership Quarterly*, 20(5), 737–748. <https://doi.org/10.1016/j.leaqua.2009.06.010>
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A Comparison of Methods to Test Mediation and Other Intervening Variable Effects. *Psychological Methods*, 7(1), 83–104. <https://doi.org/10.1037/1082-989X.7.1.83>
- Madden, E., & Ajzen, I. (1992). A Comparison of the Theory of Planned Behavior and the Theory of Reasoned Action. *Personality and Social Psychology Bulletin*, 8(1), 3–9. <https://doi.org/10.1177/01461617292181001>
- Malatji, W. R., Eck, R. V., & Zuva, T. (2020). Understanding the Usage, Modifications, Limitations and Criticisms of Technology Acceptance Model (TAM). *Advances in Science, Technology and Engineering Systems Journal*, 5(6), 113–117. <https://doi.org/10.25046/aj050612>
- Marangunić, N., & Granić, A. (2015). Technology Acceptance Model: A Literature Review from 1986 to 2013. *Universal Access in the Information Society*, 14(1), 81–95. <https://doi.org/10.1007/s10209-014-0348-1>

- Mariano, J., Marques, S., Ramos, M. R., Gerardo, F., Cunha, C. L. d., Girenko, A., Alexandersson, J., Stree, B., Lamanna, M., Lorenzatto, M., Mikkelsen, L. P., Bundgård-Jørgensen, U., Régo, S., & de Vries, H. (2022). Too Old for Technology? Stereotype Threat and Technology Use by Older Adults. *Behaviour & Information Technology*, 41(7), 1503–1514. <https://doi.org/10.1080/0144929X.2021.1882577>
- Mathieu, J. E., & Zajac, D. M. (1990). A Review and Meta-Analysis of the Antecedents, Correlates, and Consequences of Organizational Commitment. *Psychological Bulletin*, 108(2), 171.
- Matzler, K., Schwarz, E., Deutinger, N., & Harms, R. (2008). The Relationship between Transformational Leadership, Product Innovation and Performance in SMEs. *Journal of Small Business & Entrepreneurship*, 21(2), 139–151. <https://doi.org/10.1080/08276331.2008.10593418>
- McClure, P. K. (2018). "You're Fired," Says the Robot: The Rise of Automation in the Workplace, Technophobes, and Fears of Unemployment. *Social Science Computer Review*, 36(2), 139–156. <https://doi.org/10.1177/0894439317698637>
- McKinsey. (2024). The State of AI in Early 2024: Gen AI Adoption Spikes and Starts to Generate Value. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
- McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa ... Examine the Variables Influencing the Use of Artificial Intelligent In-Home Voice Assistants. *Computers in Human Behavior*, 99, 28–37. <https://doi.org/10.1016/j.chb.2019.05.009>
- Meyer, J. P., & Allen, N. J. (1991). A Three-Component Conceptualization of Organizational Commitment. *Human Resource Management Review*, 1(1), 61–89. [https://doi.org/10.1016/1053-4822\(91\)90011-Z](https://doi.org/10.1016/1053-4822(91)90011-Z)
- Mishra, D., Akman, I., & Mishra, A. (2014). Theory of Reasoned Action Application for Green Information Technology Acceptance. *Computers in Human Behavior*, 36, 29–40. <https://doi.org/10.1016/j.chb.2014.03.030>
- Mohr, S., & Kühn, R. (2021). Acceptance of Artificial Intelligence in German Agriculture: An Application of the Technology Acceptance Model and the Theory of Planned Behavior. *Precision Agriculture*, 22(6), 1816–1844. <https://doi.org/10.1007/s11119-021-09814-x>
- Molino, M., Cortese, C. G., & Ghislieri, C. (2021). Technology Acceptance and Leadership 4.0: A Quali-Quantitative Study. *International Journal of Environmental Research and Public Health*, 18(20), Article 20. <https://doi.org/10.3390/ijerph182010845>
- Morelli, N. A., & Cunningham, C. J. L. (2012). Not All Resources Are Created Equal: COR Theory, Values, and Stress. *The Journal of Psychology*, 146(4), 393–415. <https://doi.org/10.1080/00223980.2011.650734>
- Mowday, R. T., Porter, L. W., & Steers, R. M. (1982). *Employee-Organization Linkages: The Psychology of Commitment, Absenteeism, and Turnover*. Academic Press.
- Mowday, R. T., Steers, R. M., & Porter, L. W. (1979). The Measurement of Organizational Commitment. *Journal of Vocational Behavior*, 14(2), 224–247. [https://doi.org/10.1016/0001-8791\(79\)90072-1](https://doi.org/10.1016/0001-8791(79)90072-1)
- Neufeld, D. J., Dong, L., & Higgins, C. (2007). Charismatic Leadership and User Acceptance of Information Technology. *European Journal of Information Systems*, 16(4), 494–510. <https://doi.org/10.1057/palgrave.ejis.3000682>
- Nielsen, P. A., Boye, S., Holten, A.-L., Jacobsen, C. B., & Andersen, L. B. (2019). Are transformational and transactional types of leadership compatible? A two-wave study of employee motivation. *Public Administration*, 97(2), 413–428. <https://doi.org/10.1111/padm.12574>
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192. <https://doi.org/10.1126/science.adh2586>
- Nunnally, J. C. (1978). *Psychometric Theory*. McGraw-Hill.
- Ore, O., & Sposato, M. (2022). Opportunities and risks of artificial intelligence in recruitment and selection. *International Journal of Organizational Analysis*, 30(6), 1771–1782. <https://doi.org/10.1108/IJOA-07-2020-2291>
- Ozkan, S., & Kanat, I. E. (2011). e-Government adoption model based on theory of planned behavior: Empirical validation. *Government Information Quarterly*, 28(4), 503–513. <https://doi.org/10.1016/j.giq.2010.10.007>
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411–419. <https://doi.org/10.1017/S1930297500002205>
- Pawar, B. S., & Eastman, K. K. (1997). The nature and implications of contextual influences on transformational leadership: A conceptual examination. *ACADEMY OF MANAGEMENT REVIEW*, 22(1), 80–109. <https://doi.org/10.2307/259225>
- Peng, J., Li, M., Wang, Z., & Lin, Y. (2021). Transformational Leadership and Employees' Reactions to Organizational Change: Evidence From a Meta-Analysis. *The Journal of Applied Behavioral Science*, 57(3), 369–397. <https://doi.org/10.1177/0021886320920366>
- Pituch, K., & Stevens, J. (2016). *Applied multivariate statistics for the social sciences* (6th). Routledge.
- Porst, R. (2014). *Fragebogen: Ein Arbeitsbuch*. Springer Fachmedien. <https://doi.org/10.1007/978-3-658-02118-4>
- Probst, T. M., Stewart, S. M., Gruys, M. L., & Tierney, B. W. (2007). Productivity, counterproductivity and creativity: The ups and downs of job insecurity. *Journal of Occupational and Organizational Psychology*, 80(3), 479–497. <https://doi.org/10.1348/096317906X159103>
- Rafferty, A. E., & Griffin, M. A. (2004). Dimensions of transformational leadership: Conceptual and empirical extensions. *The Leadership Quarterly*, 15(3), 329–354. <https://doi.org/10.1016/j.leaqua.2004.02.009>
- Raja & Palanichamy. (2011). Leadership Styles and its Impact on Organizational Commitment. *Asia-Pacific Journal of Management Research and Innovation*, 7(3), 167–175.
- Ramli, D. Y., Soelton, M., & Psi, S. (2019). The role of transactional leadership towards organizational commitment. *International Journal of Business, Economics and Law*, 20(5), 50–57.
- Randall, D. M. (1990). The consequences of organizational commitment: Methodological investigation. *Journal of Organizational Behavior*, 11(5), 361–378. <https://doi.org/10.1002/job.4030110504>
- Ruck, K., & Welch, M. (2012). Valuing internal communication; management and employee perspectives. *Public Relations Review*, 38(2), 294–302. <https://doi.org/10.1016/j.pubrev.2011.12.016>
- Sanchez-Gomez, M., Giorgi, G., Finstad, G. L., Alessio, F., Ariza-Montes, A., Arcangeli, G., & Mucci, N. (2021). Economic Stress at Work: Its Impact over Absenteeism and Innovation. *International Journal of Environmental Research and Public Health*, 18(10), Article 10. <https://doi.org/10.3390/ijerph18105265>
- Sarkar, S., & Osiyevskyy, O. (2018). Organizational change and rigidity during crisis: A review of the paradox. *European Management Journal*, 36(1), 47–58. <https://doi.org/10.1016/j.emj.2017.03.007>
- Schepers, J., Wetzels, M., & de Ruyter, K. (2005). Leadership styles in technology acceptance: Do followers practice what leaders preach? *Managing Service Quality: An International Journal*, 15(6), 496–508. <https://doi.org/10.1108/09604520510633998>
- Schultz, R. L., & Slevin, D. P. (1975). *Implementing operations research/management science*. American Elsevier Pub. Co.
- Schyns, B., & Day, D. (2010). Critique and review of leader-member exchange theory: Issues of agreement, consensus, and excellence. *European Journal of Work and Organizational Psychology*, 19(1), 1–29. <https://doi.org/10.1080/13594320903024922>
- Sheatsley, P. B. (1983). Questionnaire Construction and Item Writing. In *Handbook of Survey Research*. Academic Press Inc.
- Sheppard, B. H., Hartwick, J., & Warshaw, P. R. (1988). The Theory of Reasoned Action: A Meta-Analysis of Past Research with Recommendations for Modifications and Future Research. *Journal of Consumer Research*, 15(3), 325–343. <https://doi.org/10.1086/209170>
- Shih, Y., & Fang, K. (2004). The use of a decomposed theory of planned behavior to study Internet banking in Taiwan. *Internet Research*, 14(3), 213–223. <https://doi.org/10.1108/10662240410542643>
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7(4), 422–445. <https://doi.org/10.1037/1082-989X.7.4.422>
- Smith, A., & Anderson, J. (2014). AI, Robotics, and the Future of Jobs. *Pew Research Center*, 6.

- Srivastava, R., Shneikat, B., Mendoza, S., Elrehail, H., & Afifi, M. A. (2024). Assessing AI Adoption in the Workplace Through the Theory of Planned Behavior. *2024 2nd International Conference on Cyber Resilience (ICCR)*, 1–4. <https://doi.org/10.1109/ICCR61006.2024.10533108>
- Staw, B. M., Sandelands, L. E., & Dutton, J. E. (1981). Threat Rigidity Effects in Organizational Behavior: A Multilevel Analysis. *Administrative Science Quarterly*, 26(4), 501–524. <https://doi.org/10.2307/2392337>
- Thibaut, J. W. (1959). *The Social Psychology of Groups*. Routledge. <https://doi.org/10.4324/9781315135007>
- Toh, S.-Y., Ng, S.-A., & Phoon, S.-T. (2023). Accentuating technology acceptance among academicians: A conservation of resource perspective in the Malaysian context. *Education and Information Technologies*, 28(3), 2529–2545. <https://doi.org/10.1007/s10639-022-11288-x>
- Tourish, D., & Pinnington, A. (2002). Transformational Leadership, Corporate Cultism and the Spirituality Paradigm: An Unholy Trinity in the Workplace? *Human Relations*, 55(2), 147–172. <https://doi.org/10.1177/0018726702055002181>
- Van Hooft, A., Niesen, W., & De Witte, H. (2019). Does job insecurity hinder innovative work behaviour? A threat rigidity perspective. *Creativity and Innovation Management*, 28(1), 19–29. <https://doi.org/10.1111/caim.12271>
- van Knippenberg, D., & Sitkin, S. B. (2013). A Critical Assessment of Charismatic—Transformational Leadership Research: Back to the Drawing Board? *Academy of Management Annals*, 7(1), 1–60. <https://doi.org/10.5465/19416520.2013.759433>
- Vander Elst, T., De Witte, H., & De Cuyper, N. (2014). The Job Insecurity Scale: A psychometric evaluation across five European countries. *European Journal of Work and Organizational Psychology*, 23(3), 364–380. <https://doi.org/10.1080/1359432X.2012.745989>
- Vella, J., Caruana, A., & Pitt, L. (2013). Organizational commitment and users' perception of ease of use: A study among bank managers. *Journal of Management Development*, 32(4), 351–362. <https://doi.org/10.1108/02621711311326356>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, 36(1), 157–178. <https://doi.org/10.2307/41410412>
- Xerri, M. J., & Brunetto, Y. (2013). Fostering innovative behaviour: The importance of employee commitment and organisational citizenship behaviour. *The International Journal of Human Resource Management*, 24(16), 3163–3177. <https://doi.org/10.1080/09585192.2013.775033>
- Yousafzai, S. Y., Foxall, G. R., & Pallister, J. G. (2010). Explaining Internet Banking Behavior: Theory of Reasoned Action, Theory of Planned Behavior, or Technology Acceptance Model? *Journal of Applied Social Psychology*, 40(5), 1172–1202. <https://doi.org/10.1111/j.1559-1816.2010.00615.x>
- Yousef, D. A. (2000). Organizational commitment: A mediator of the relationships of leadership behavior with job satisfaction and performance in a non-western country. *Journal of Managerial Psychology*, 15(1), 6–24. <https://doi.org/10.1108/02683940010305270>