

**Advisory Editorial Board:**

DOMINIK VAN AAKEN  
 FREDERIK AHLEMANN  
 CHRISTOPH BODE  
 ROLF BRÜHL  
 JOACHIM BÜSCHKEN  
 LEONHARD DOBUSCH  
 RALF ELSAS  
 DAVID FLORYSIK  
 GUNTHER FRIEDL  
 WOLFGANG GÜTTEL  
 CHRISTIAN HOFMANN  
 KATJA HUTTER  
 LUTZ JOHANNING  
 STEPHAN KAISER  
 ALFRED KIESER  
 NATALIA KLIEWER  
 DODO ZU KNYPHAUSEN-AUFSEß  
 SABINE T. KÖSZEGI  
 ARJAN KOZICA  
 TOBIAS KRETSCHMER  
 HANS-ULRICH KÜPPER  
 REINER LEIDL  
 ANTON MEYER  
 MICHAEL MEYER  
 GORDON MÜLLER-SEITZ  
 J. PETER MURMANN  
 BURKHARD PEDELL  
 MARCEL PROKOPCZUK  
 TANJA RABL  
 SASCHA RAITHEL  
 ASTRID REICHEL  
 KATJA ROST  
 MARKO SARSTEDT  
 DEBORAH SCHANZ  
 ANDREAS G. SCHERER  
 STEFAN SCHMID  
 UTE SCHMIEL  
 CHRISTIAN SCHMITZ  
 PHILIPP SCHRECK  
 GEORG SCHREYÖGG  
 LARS SCHWEIZER  
 DAVID SEIDL  
 THORSTEN SELLHORN  
 ANDREAS SUCHANEK  
 ORESTIS TERZIDIS  
 ANJA TUSCHKE  
 SABINE URNIK  
 STEPHAN WAGNER  
 BARBARA E. WEIßENBERGER  
 ISABELL M. WELPE  
 HANNES WINNER  
 CLAUDIA B. WÖHLE  
 THOMAS WRONA  
 THOMAS ZWICK

# JUNIOR MANAGEMENT SCIENCE

|                                                                                                                                                |     |
|------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Carina Brettschneider</b> , R&D Accounting Discretion as an Income Smoothing Tool: An Empirical Analysis of German Listed Companies         | 151 |
| <b>Theresa Langenmayr</b> , Implementation of Strategic Change Franchisees: A Sensemaking Perspective                                          | 173 |
| <b>Maximilian Bölstler</b> , The Influence of Cryptocurrencies on Enterprise Risk Management – an Empirical Evidence by the Example of Bitcoin | 195 |
| <b>Daniel Schürk</b> , Does Subordinated Debt Discipline Banks? Empirical Evidence of Market Discipline in Europe                              | 228 |
| <b>Philipp Küst</b> , The Impact of the Organic Label Halo Effect on Consumers' Quality Perceptions, Value-in-Use and Well-Being               | 241 |
| <b>Tom O. K. Zeissler</b> , Charakteristika vs. Carry - Outperformance in Devisenmärkte                                                        | 265 |



# R&D Accounting Discretion as an Income Smoothing Tool: An Empirical Analysis of German Listed Companies

Carina Brettschneider

*Ludwig-Maximilians-Universität München*

## Abstract

Capitalization of development costs is compulsory according to IFRS if a set of criteria is fulfilled. However, this obligation is considered as a de facto right for capitalization since the criteria are quite subjective, allowing for a certain degree of flexibility. Hence, the question arises whether managers use research and development (R&D) accounting to conduct earnings management in terms of income smoothing. Using a sample of German listed companies, the study conducts several regression analyses to test whether there is a negative relationship between R&D capitalization and different income smoothing proxies. Results show that the hypothesis is supported independent of the income smoothing proxy used. The study proves that managers indeed use R&D capitalization judgments to conduct income smoothing.

**Keywords:** R&D capitalization; income smoothing; earnings management; R&D accounting; development costs.

## 1. Introduction

“Nature builds no machines, no locomotives, railways [...] etc. [...]. They are organs of the human brain, created by the human hand; the power of knowledge, objectified” (Marx, 1857).

What we currently observe as the transformation from a manufacturing to a knowledge-based economy has already been envisaged by Karl Marx in 1857. Digitalization and globalization led to a technology-caused structural interruption that puts substantial pressure on current business models and whole industries. This forces companies to heavily invest in research and development (R&D) to remain competitive and improve their businesses (Hoffmann, 2017). As reported by the R&D Magazine’s current 2017 Global R&D Funding Forecast, worldwide spending on R&D reaches \$ 2.066 trillion<sup>1</sup>, which represents 1.72 % of the global gross domestic product (Industrial Research Institute, 2017).

This ongoing development has also been recognized by German politics. The new coalition agreement (2018) intends to introduce tax incentives to promote R&D investments. Additionally, 3.5 % of gross domestic product are planned to spend on R&D until 2025 to support high-tech strategies such as artificial intelligence (Koalitionsvertrag, 2018).

Obviously, the transformation towards a R&D intensive economy is also reflected in the companies’ financial statements. According to a study conducted by PricewaterhouseCoopers GmbH in 2014, intangible assets determine 75 % of a firm’s value (Hadjiloucas, 2014). Simultaneously, several researchers denounce exhaustive sluggishness in R&D measuring and reporting systems (Boulton et al., 2000). Neither literature nor standard setters can agree on a uniform accounting treatment leading to a variety of different rules among countries (Garanina et al., 2016). While some presume R&D capitalization to be value relevant signaling future economic benefits, opponents condemn the loss of objectivity and the possible incentive for earnings management, demanding complete expensing of R&D (Lev and Sougiannis, 1996).

Following this debate, the thesis wants to assess whether managers really use R&D capitalization opportunities to engage in income smoothing in terms of accrual-based earnings management. Since research on that topic does not lead to a consistent result, the thesis contributes to previous literature by conducting an analysis in a setting of German listed companies over a period of four years and across different industries.

Section two introduces R&D accounting in Germany focusing mainly on International Financial Reporting Standards (IFRS). After discussing the trade-off between objec-

<sup>1</sup>Hereinafter, the thesis uses German notation of numbers.

tivity and relevance, focus shifts on earnings management research and the income smoothing hypothesis. Whereas section four reviews previous literature, the following chapter develops the research hypothesis analyzed. Section six and seven are concerned with research methodology and the empirical data analysis that consists of a descriptive statistic, correlation, regression and sensitivity analysis. In the end, the thesis critically assesses the analysis made by pointing out limitations and further research opportunities before drawing a conclusion.

## 2. R&D accounting in Germany

### 2.1. Terminology, recognition and measurement

Introducing the Accounting Law Modernization Act in Germany in 2009, legislature lifts the ban on capitalization of self-generated intangibles assets as stated in § 248 (2) German Commercial Code (HGB). Companies got the option to recognize development expenses at acquisition costs while research expenses are not allowed to be capitalized. According to HGB, development costs mainly include applied research results or knowledge for further advancement or recent developments, while research focuses on the search for knowledge itself. If companies are not able to separate research and development, they must account both as an expense (§ 255 (2a) HGB).

Since European International Accounting Standards (IAS) regulation article 4 requires capital market orientated companies to prepare their consolidated financial statements in accordance to IFRS starting in 2005, it is not appropriate to apply HGB when analyzing German listed companies. Therefore, the thesis focuses on IFRS accounting principles. Accounting for intangible assets including R&D is governed in IAS 38 (Figure 1). In line with German legislation, IFRS distinguishes between research and development phase. Although the definitions are quite similar, they are treated differently. During the research phase, the company is not able to demonstrate an intangible asset that is expected to create future economic benefits. Since recognition criteria for an intangible asset pursuant to IAS 38.21 are not fulfilled in the research phase, these costs must be charged as an expense. This mainly includes obtaining knowledge, evaluating research or selecting alternatives (IAS 38.54 - 56).

In the development phase, expenditures are obligated to be recognized as an intangible asset if the basic definition and recognition principles (IAS 38.8 – 28.23) in addition to the following six criteria are satisfied:

1. Technical feasibility
2. Intention of completion for sale or use
3. Ability to use or sale
4. Proof of how to generate future economic benefits
5. Availability of sufficient resources (technical, financial and other) to complete the development
6. Reliability in measuring attributable costs

Design, construction and testing of alternatives, pre-production or pilot plants can be considered as appropriate examples (IAS 38.57 - 60). According to IAS 38.71 restatement of past expenses is prohibited. Additionally, IFRS forces companies to expense their R&D if they are not able to distinguish between research and development phase (IAS 38.53).

If all criteria for recognition are fulfilled, R&D expenditures that arose from that date on are capitalized at cost. These expenditures include all directly attributable costs like materials, employee benefits, fees etc. to bring the asset into the intended shape (IAS 38.65 – 38.67). For subsequent measurement, IAS 38.72 requires the use of the cost model unless there is an active market. If the latter applies, companies get the option to adopt the revaluation model.

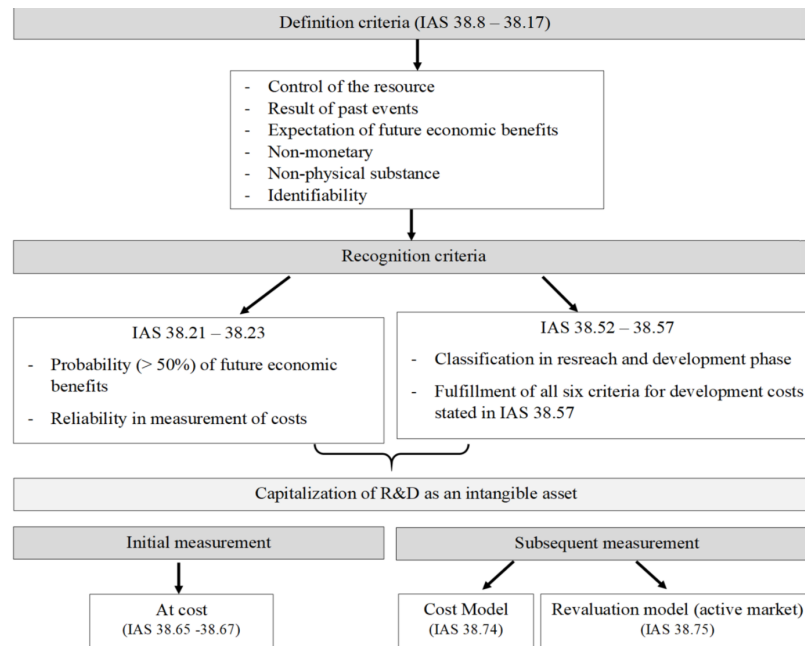
In addition, the entity has to determine whether the useful life of the asset capitalized is finite or infinite using the criteria stated in IAS 38.90. While a finite useful life calls for amortization, an infinite one requires the company to conduct an impairment test in line with IAS 36 on an annual basis and if an indication for impairment arises (IAS 38.97; IAS 38.108).

In terms of disclosure, the company has to differentiate between internally generated and other intangible assets. These categories need to include gross carrying amount along with a reconciliation at the end and beginning of the period, useful life (if any), amortization method and accumulated amount including impairment, reasons supporting various assessments and other special requirements stated in IAS 38.118 – 38.128. Although IFRS capitalization policy differs from German standards as they only provide an option for capitalization, the discrepancy is rather formal than substantial. The mentioned criteria obligating capitalization of R&D are considered as quite subjective and therefore allow a certain degree of flexibility. If managers intend to expense some of their R&D instead of capitalizing, they can simply justify the assessment. Therefore, R&D accounting might be driven by earnings management incentives such as to smooth income (Markarian et al., 2008).

### 2.2. The trade-off between relevance and objectivity

Hardly any topic shows as much different accounting treatments among countries as R&D (Garanina et al., 2016). The decision to allow or even force companies to capitalize R&D expenditures is mainly a trade-off between relevance and objectivity. Financial Accounting Standard Board (FASB) forces US companies to expense all their R&D costs. Capitalization is prohibited due to the fear of impaired objectivity. FASB argues that expensing prevents managers from capitalizing R&D projects that show low probability to create future economic benefits. Additionally, expensing should avoid managers to conduct earnings management when deciding to capitalize R&D. (Cazavan-Jeny et al., 2011; Healy et al., 2002; Lev and Sougiannis, 1996).

International Accounting Standard Board (IASB) as well as German legislature, focus on the value relevance aspect by allowing capitalization. Thereby they respond to the arising importance of R&D, arguing that it represents a high value



**Figure 1:** R&D accounting according to IFRS (own figure based on IAS 38)

asset whose disregard would earnestly decrease credibility and relevance of financial statements (Healy et al., 2002). It is reasoned that R&D creates a positive impact on profitability and market value and thus provides value relevant information to shareholders and investors (Lev and Sougiannis, 1996; Sougiannis, 1994).

### 3. Earnings management through R&D

#### 3.1. Definition and incentives

The most popular definition of earnings management is given by Schipper (1989), indicating that it is “the purposeful intervention in the external financial reporting process, with the intent of obtaining some private gain”. Since financial reports should reflect managers’ perception of the firms’ performance, accounting standards have to allow for a certain range of judgment. However, managers can also use this flexibility to alter reports in a way that deceives stakeholders about the true economic performance or influences contractual outcomes that are based on accounting numbers (Healy and Wahlen, 1999). This issue exactly reflects the problem of discretion in R&D accounting. On the one hand, standard setters of the IASB want to enable a certain degree of judgment since managers themselves are the best qualified to assess whether a R&D project will create future economic benefits and therefore should be recognized as an asset. On the other hand, it is obvious that managers might tend to act opportunistically to improve outcomes and thus misuse the given flexibility. However, it is important to distinguish earnings management from fraud. In contrast to fraud, earnings management uses accounting judgements that are in line with legal boundaries and generally accepted accounting principles (Makar et al., 2000).

Earnings management through R&D capitalization represents a kind of accrual-based earnings management since managers defer expenditures to subsequent periods when the intangible asset will be amortized (if useful life is finite) or impaired. Thus, managers can deal with accruals to improve their results. The total amount spent on R&D is a way to conduct real earnings management as the amount of R&D investments directly influences financial statements (Enomoto et al., 2015).

Incentives to exercise earnings management through R&D comprise income smoothing, bonus plans and debt covenants (Markarian et al., 2008). The latter proposes that managers are more likely to choose income increasing accounting policies when their companies are high leveraged and closer to violate debt covenants (Duke and Hunt III, 1990). The bonus plan incentive suggests that managers tend to choose accounting policies that improve their bonuses (Healy and Wahlen, 1999). However, the focus of the thesis will be on income smoothing caused by R&D capitalization, meaning that managers apply earnings management to reduce income fluctuations (Hepworth, 1953). This is justified by the lack of sufficient data to analyze the bonus plan hypothesis and by the restricted extent of the thesis to include the debt covenant incentive. Additionally, income smoothing is the most examined hypothesis in the current literature regarding R&D accounting discretion.

#### 3.2. The income smoothing hypothesis

Income smoothing comprises all tools managers may use in order to scale down fluctuations of disclosed income figures, with regard to some income targets, using artificial (accounting) or real (transactional) variables (Koch, 1981; Leuz et al., 2003). There are several accounting techniques to



smooth income like gross revenue manipulation, overproduction of finished goods, property accounting or intangible asset accounting, which also includes R&D (Hepworth, 1953). As this thesis focuses on accounting discretion, only artificial variables are considered, looking at how managers potentially use accruals in terms of R&D capitalization, to reduce income variability. Concerning the income smoothing hypothesis, income reflects the normal earnings power of a company, which is reported as operating income or earnings before interest and tax (EBIT) in the companies' financial statements (Koch, 1981). This is consistent with prior studies indicating operating profitability as the income smoothing target (Barnea et al., 1976; Makar et al., 2000; Moses, 1987). Therefore, in the following, earnings and income are used synonymously. However, a sensitivity analysis will be conducted later in section 7.4. to examine if there are probably other income smoothing targets such as net income or earnings per share (EPS) since it cannot be presumed that operating profitability is the company's only smoothing target (Imhoff, 1981).

Motives for managers to smooth earnings over time through R&D are multi-layered. One of the most important aspects is the reduced risk through low income fluctuation. The lower the observed volatility, the lower is the perceived assessment of various claimants about the stability and potential bankruptcy of the company. Thus, borrowing costs of the enterprise decrease as debt capital providers are more confident in getting back their money on time (Trueman and Titman, 1988). Simultaneously, the selling price of shares increases as (risk averse) investors tend to prefer companies characterized by a low risk. Stable earnings are also desirable from a shareholder's perspective since they are better suited to indicate higher dividend payment than variable earnings. (Beidleman, 1973). Other stakeholders also feel more confident with an income smoothing approach, as it reflects security and stability, leading to higher overall satisfaction (Hepworth, 1953).

Additionally, the company itself benefits from a smooth income. Having huge fluctuation in earnings, it becomes difficult to develop detailed plans and budgets for further periods, as these are based on past performance. Furthermore, managers themselves are assessed by budgets established (Beidleman, 1973). Since volatility in the firm's income could be a potential reason to displace managers, they are interested in smooth income. Also income based compensation might serve as an incentive for managers to conduct income smoothing (Brayshaw and Eldin, 1989).

Smooth income further incorporates tax advantages. In case of a progressive tax rate for example, companies would show direct tax savings by smoothing their income, compared to high volatility (Hepworth, 1953).

The underlying reason to engage in income smoothing determines whether the managers' aim is to convey private, internal information to stakeholders like described by Beidleman (1973) or to reach some opportunistically driven targets (Vander Bauwhede et al., 2003). It is important to consider that income smoothing also faces some difficulties. Assessing

the right amount that has to be adjusted is often quite ambitious. It involves guesses and estimations that might lead to over- or under adjustment and thus counter-smoothing. (Vander Bauwhede et al., 2003).

#### 4. Literature review

Starting in the mid-20th century, earnings management has been developed as one of the main research areas in accounting literature. First, the focus was primarily on detecting and investigating earnings management followed by an analysis of incentives, conditions and tools managers might use to alter earnings. In 1953, Hepworth was one of the first researchers who looked at possible motivations to engage in income smoothing as a part of earnings management. At that time, various research evolved, indicating that manager use accounting discretion and choices to influence earnings and thus proofed the existence of earnings management. Burgstahler and Dichev (1997) detected that there is an unusual high amount of slightly positive earnings compared to small negative ones, giving the evidence that companies engage in earnings management to avoid reporting losses. The income-smoothing hypothesis was confirmed i. a. by Beidleman (1973) and Koch (1981), who indicated that income smoothing is higher if costs for it are low and ownership is diverse.

As R&D investments have become more and more important in the early 2000s, researchers were concerned with the best R&D accounting policy that avoids discretion and therefore earnings management. Various studies as those performed by Lev and Sougiannis (1996) or Zhao (2002) proofed that R&D capitalization is value relevant and should therefore be the preferred accounting method. Cazavan-Jeny and Jeanjean (2003) described R&D capitalization as a signal for stakeholders as it is positively related with stock returns. This was supported by Oswald and Zarowin (2007) who showed that the association between current year returns and future earnings is higher in case of R&D capitalization. However, Chan et al. (2007) claimed that higher R&D intensity indicates company's performance, independent of the accounting method used. Wang et al. (2017) considered R&D accounting policy mainly as a trade-off between accounting performance that is supported through expensing and market value that is increased by capitalization.

Real earnings management was examined by Osma and Young (2009), Cazavan-Jeny et al. (2011) and Tahinakis (2014), who concluded that companies cut their R&D investments to avoid reporting losses or to show positive earnings. Additionally, Seybert (2010) proofed that managers tend to overinvest in R&D in case of capitalization due to reputation concerns.

Besides, managers might also use accrual-based earnings management through R&D capitalization. However, research on that topic is relatively young and result are inconsistent. Zicke (2014) and Abrahams and Sidhu (1998) came to the result that managers capitalize more of their R&D expenditures to avoid reporting losses or earnings decrease. Further-

more, Markarian et al. (2008) and Triki-Damak and Halioui (2013) examined that managers use R&D capitalization to smooth earnings over periods in a setting of French and Italian companies. This is only partly supported by Persson and Fuentes (2011) who could merely indicate income smoothing through R&D capitalization in one of three periods looking at Swedish enterprises. This difference can be explained by Garanina et al. (2016) who indicated that earnings management incentives to capitalize R&D vary across countries. While income smoothing through R&D was detected in Germany, Russia used R&D to meet debt covenants which was also proofed in France by Triki-Damak and Halioui (2013). However, the study conducted by Guidara et al. (2014) in France suggested that managers do not use R&D accounting discretion to smooth earnings but rather support the real earnings management literature.

## 5. Hypothesis development

As already stated in section 2.1., R&D accounting according to IFRS allows for a certain degree of flexibility. The de facto right for capitalization enables managers to easily decide how to account for R&D based on the current economic situation. If they prefer expensing instead of capitalization or vice versa they can simply substantiate this approach by arguing that they are not able to generate future economic benefits or costs on a project cannot be measured reliable for capitalization (Garanina et al., 2016; Makar et al., 2000; Markarian et al., 2008).

The discretion in accounting for R&D given by IFRS can be used by managers as an instrument to conduct income smoothing in terms of accrual-based earnings management as described in section 3. Auditors confirm that accruals are among the most common methods to manage earnings. As R&D capitalization choice is based on accruals by deferring development expense to future periods through amortization/impairment, it is reasonable to argue that managers use R&D accounting to engage in earnings management to smooth income (Nelson et al., 2003; Persson and Fuentes, 2011). There are many studies indicating that managers do smooth their income and have reasonable incentives (see 3.2. and 4.) such as to reduce borrowing cost or increase share prices (Beidleman, 1973; Trueman and Titman, 1988).

To moderate fluctuation in their income managers first analyze their pre-managed operating profitability as this is presumed to be the income smoothing target (Barnea et al., 1976; Makar et al., 2000; Moses, 1987). If they realize that the current pre-managed operating profitability is below the previous years' profitability measure, they opt for accounting methods that increase reported earnings and vice versa (Markarian et al., 2008). As R&D accounting policy directly influences the extend of earnings, managers can control profitability by choosing to expense or capitalize development costs (Persson and Fuentes, 2011).

Thus, it is expectable that managers prefer to capitalize a higher magnitude of their R&D expenditures if their pre-managed operating profitability before R&D capitalization is

below the profitability of previous years. In return, expensing is observed when pre-managed operating profitability before R&D capitalization is above the past years' profitability. Before R&D capitalization means that the amount of development costs recorded as an intangible asset is not considered in the pre-managed profitability, assuming that all costs are expensed first. Capitalization is determined only after comparing the pre-managed profitability to previous years (Markarian et al., 2008). Hence, managers decide on the amount of development costs capitalized based on the company's profitability change with the aim to report a smooth income over the periods (Persson and Fuentes, 2011). Since profitability commonly defines the relation between a company's profit and the size of the business, this analysis defines the change in operating profitability as the current year's pre-managed EBIT minus the average of the last two years' EBIT normalized by the current year's total assets before R&D capitalization (Markarian et al., 2008). This aspect will be explained in section 6.2.1.

Therefore, the hypothesis examined is formulated as follows:

H: There is a negative relationship between the change in a firm's operating profitability and R&D capitalization.

## 6. Research methodology

### 6.1. Sample selection

Data for the subsequent empirical analysis is gathered from German listed companies. Germany is regarded to be a suitable object of research because it is known to be one of the innovation leaders, showing the highest R&D investments in Europe in 2016 (Hernández et al., 2017).

At first, all German listed companies are assigned to their respective sectors (sector and industry are considered interchangeably) based on Global Industry Classification Standard (GICS). Additionally, the category "conglomerate" is added for companies engaging in various fields. Financials, real estate, energy and consumer staples sectors are excluded from the analysis since they do not show any or sufficient R&D investments or do not publish them and are therefore not appropriate to be considered.

Secondly, companies within an industry are ranked based on their market capitalization. Within each industry annual reports are analyzed from the greatest to the smallest company over a period of four years (2013 – 2016). Those who do not show any R&D capitalization, do not disclose them concretely or refuse to disaggregate capitalized development cost from other intangible asset are removed from the list. Still, it cannot be assumed that these companies do not engage in earnings management through R&D (Persson and Fuentes, 2011).

For the remaining companies, the largest of each sector are incorporated to end up with a total sample size of 92. Recapitulated, the analysis is based on an entire population of 23 companies over a time range of four years. Annual

reports of six years have to be analyzed for each company since the estimation of changes over periods requires to go back further in the past for the first year incorporated (see 6.2.1.).

For the empirical analysis, R&D numbers have to be collected manually, looking at the annual reports of the companies. R&D capitalization can either be disclosed directly within the consolidated balance sheet, in the intangible asset movement schedule or in the notes. Other numbers necessary were partially gathered from the data platform WRDS. For the remaining individual figures and ratios, formulas were calculated with Excel. To conduct the empirical analysis, SPSS is used as a statistical software.

Data selection reveals that the extend of R&D disclosure is quite different. While the amount of qualitative information is very extensive this is not always the case for quantitative reports. Sometimes it is difficult to determine the exact amount capitalized during the year as companies hesitate to disclose a large scale of figures. However, this has been improving over the years, showing that R&D has been becoming more and more important for companies to communicate to their stakeholders. Approximately every fourth euro expensed for R&D currently reappears as an asset in the companies' balance sheets but the amount capitalized differs across industries (Leibfried & Pfanzelt, 2014).

Capitalization in the pharmaceutical industry is rather rare although it is known to be R&D intensive. Due to the high risks until the pharmaceutical product approval is reached, companies are forced to expense R&D. The most common reason refraining from capitalization is the lack of probable future economic benefits.

## 6.2. Specification of variables

### 6.2.1. Dependent and independent variables

For each observation of the sample, a set of variables must be defined and calculated in order to perform the empirical analysis. First, RDCapitalization is introduced as the dependent variable. It is calculated as the absolute gross amount of development cost capitalized in each year divided by the company's total assets (Markarian et al., 2008; Persson and Fuentes, 2011).

Amortization and impairment on intangible assets of the respective year are not considered in the total amount capitalized (as some prior studies did) because they are not subjected to any accounting choice and additionally could also refer to previous years' issues. Another approach would be to include the amortized and impaired amount of the development cost capitalized during the respective year, if there are any, since they reappear in the income statement and thus reduce earnings. This would lead to an opposite effect reducing the income increasing consequence of R&D capitalization, but this information is not published in any annual report. Additionally, these amounts would be very small and therefore are not presumed to change the results significantly. Consistent with preceding earnings management studies, total assets serve as a deflator (Jones, 1991). Another way to

scale R&D capitalization could be net income as it is directly related to earnings. This approach could bias the results since all years which show negative values would have to be removed from the sample (Markarian et al., 2008).

To examine the established hypothesis above,  $\Delta RoA$  serves as the independent variable. It reflects the change in return on assets (RoA) over the average of two years and represents the operating profitability progress of the company. Thus,  $\Delta RoA$  is calculated by subtracting the average of the last two years' EBIT from the current year's pre-managed EBIT (without R&D capitalization) normalized by the current year's total assets before R&D capitalization (Markarian et al., 2008).

$$\Delta RoA_t = \frac{(EBIT_t - \text{R\&D capitalized}_t) - \frac{EBIT_{t-1} + EBIT_{t-2}}{2}}{\text{Total assets}_t - \text{R\&D capitalized}_t} \quad (1)$$

Since R&D capitalized is deducted from current EBIT, it is appropriate to equally adjust current assets by subtracting the same amount. How to exactly account for  $\Delta RoA$  slightly differs in previous studies. It is not clearly defined whether total assets are used with or without R&D capitalization. (Garanina et al., 2016; Triki-Damak and Halioui, 2013; Markarian et al., 2008; Persson and Fuentes, 2011). However, as the amount capitalized during a year is relatively low compared to total assets, this should not lead to profound changes in the results.

The average over the previous two fiscal years is used because it is assumed that managers as well as the market consider the change in operating profitability as the change in "base" earnings, rather than the change related to the latest earnings reported (Markarian et al., 2008). This approach has been examined by Markarian et al. (2008) who indicate that  $R^2$  decreases when taking the change over one instead of two periods. This aspect will also be examined in section 7.4.

Even though this thesis solely focuses on R&D capitalization to smooth income, it could not be ruled out that managers simultaneously use other accruals or operating decisions to conduct earnings management. Thus, it could not be assumed that pre-managed RoA is not influenced by other instruments (Markarian et al., 2008; Persson and Fuentes, 2011).

### 6.2.2. Control variables

To decrease variance error and increase validity of the empirical analysis, several control variables are added to the model. The first one comprises the current year's pre-managed RoA, which is calculated as the company's EBIT before R&D capitalization divided by current total assets less capitalized R&D. The present operating profitability serves as a proxy for the expected future operating profitability since this data is not available (Miller and Skinner, 1998). Expected future operating profitability is presumed to be an indicator for successful completion of R&D projects. A high value increases the probability to obtain future economic benefits (one of the six criteria stated in IAS 38 for

**Table 1:** Sample by industries (own table)

| Industry                   | Number of companies | Number of observations | Percentage |
|----------------------------|---------------------|------------------------|------------|
| Conglomerate               | 2                   | 8                      | 9%         |
| Consumer discretionary     | 4                   | 16                     | 17%        |
| Health care                | 4                   | 16                     | 17%        |
| Industrials                | 4                   | 16                     | 17%        |
| Information technology     | 2                   | 8                      | 9%         |
| Materials                  | 3                   | 12                     | 13%        |
| Telecommunication services | 2                   | 8                      | 9%         |
| Utilities                  | 2                   | 8                      | 9%         |
| Total                      | 23                  | 92                     | 100%       |

R&D capitalization). Further, current profitability might also signal that the company commands sufficient resources to finalize and use the R&D project, which also reflects one of the six capitalization criteria according to IFRS. Therefore, expected future profitability and R&D capitalization might be positively related (Markarian et al., 2008). Though, the argumentation also works the other way around considering the earnings management approach. Profitable firms have an incentive to expense large parts of their R&D in order to decrease tax payments as well as political costs. Companies characterized by low profitability tend to capitalize as much of their R&D as possible to eke out achievements (Markarian et al., 2008; Persson and Fuentes, 2011). Additionally, Cazavan-Jeny and Jeanjean (2003) argue that highly profitable companies prefer to expense their R&D as they do not want to ruin the analysts' image concerning their earnings quality. This is also supported by Aboody and Lev (1998) who examined that less profitable firms capitalize more of their development costs than high profitable ones. Following this argumentation, a negative relationship should be observed between RDCapitalization and RoA. Hence, from a methodological point of view, the association cannot be clearly determined in advance.

RDTotal represents the second control variable. It comprises the total amount of R&D investments spent by the company within one year divided by total assets in the current fiscal year before R&D capitalization. It could be assumed that more R&D projects fulfill IAS 38's recognition criteria if the companies' R&D investments are higher. Thus, managers capitalize more in case of higher R&D spending, indicating a positive relation between RDTotal and RDCapitalization. Nevertheless, IAS 38 incorporates a certain degree of subjectivity, meaning that managers can decide on how much to capitalize by arguing appropriately. In addition, it has to be considered that there is a variety of other aspects and circumstances that determine managers' R&D accounting policy such as conservative accounting behavior (Garanina et al., 2016; Markarian et al., 2008). Besides, there are arguments indicating a negative association between RDTotal and RDCapitalization. R&D intensive enterprises used to carry out more R&D projects than companies spending less on R&D. Hence, it is more complex and elaborate for

them to review whether recognition criteria are fulfilled for each single project. Since this is perceived to be very time consuming, managers might simply expense R&D (Garanina et al., 2016; Markarian et al., 2008). It could also be argued that companies which spend less on R&D in a period (maybe because they are facing bad times) rather continue to work on existing projects. Thus, the probability of those projects to fulfill recognition criteria increases. In consequence, it is not possible to make ex ante predictions concerning the relation of RDTotal and RDCapitalization (Garanina et al., 2016; Markarian et al., 2008).

Given that high leveraged companies are presumed to have an incentive to capitalize more of their R&D costs to improve their result, (financial) Leverage is introduced as third control variable. It is computed as the fraction of total debt from total assets before R&D capitalization (Markarian et al., 2008). Managers that avoid violating debt covenants through R&D capitalization increase their probability to get attractive loans and reduce (perceived) risk of the business. Therefore, a positive relationship between Leverage and RDCapitalization is expected (Aboody and Lev, 1998; Cazavan-Jeny and Jeanjean, 2003; Hamada, 1972).

(Firm) Size is defined as the natural logarithm of current year's total assets before R&D capitalization effects (Markarian et al., 2008). Due to nature of research this variable is implemented to control for the companies' political sensitivity and market/political visibility as these aspects seem to influence accounting decisions of managers and thus R&D capitalization (Liberty and Zimmerman, 1986; Markarian et al., 2008). Since larger companies tend to be more sensitive to political cost pressure, they are exposed to higher capital transfers. Therefore, governments focus more on huge firms as their ability to contribute to state income is higher. Companies want to avoid this pressure by reducing reported earnings. In terms of R&D, managers will prefer expensing as it decreases the company's income (Aboody and Lev, 1998; Dufour and Zemzem, 2005; Triki-Damak and Halioui, 2013; Othman and Zéghal, 2006). But not only the government pays particular attention to large companies. As the company size increases, managers face more and stronger examination by several analysts. The growing visibility rather prevents managers from excessively influencing reported earn-



ings (Garanina et al., 2016; Othman and Zéghal, 2006). Another aspect why to consider size effects has been indicated by Cazavan-Jeny and Jeanjean (2003). They propose that large enterprises primarily invest in basic research, maintenance and product upgrades. Major parts of these costs typically do not fulfill recognition criteria of IAS 38 and therefore must be expensed. These suggestions anticipate a negative relationship between Size and RDCapitalization.

High-Capitalizer as the fourth control variable is calculated as a dummy variable equal to one if the capitalized R&D amount normalized by earnings is above the median of the sample firms and zero if it is below the median. High-Capitalizer is introduced to control for a potential, significant impact of R&D capitalization on earnings (Markarian et al., 2008).

LagCapitalization as the fifth variable refers to capitalized R&D during the previous year, since lagged variables are considered as a strategy to control for endogeneity (Boone et al., 2007). The decision not to take LagCapitalization into account excludes a seriously correlated variable (Markarian et al., 2008). Further, the variable implies how consistent companies perform their R&D accounting policy (Zicke, 2014). However, it is quite difficult to determine a company's consistency if managers conduct earnings management. Additionally, it has to be kept in mind that R&D investments are basically difficult to predict and outcomes of R&D spending could fluctuate over the years (Garanina et al., 2016). Even though there might be concerns due to serial correlation between LagCapitalization and RDCapitalization, Markarian et al. (2008) proofed that worries are unfounded using Durbin Watson statistic and calculating a first order autocorrelation.

To control for future growth opportunities, (stock market) performance and risk of a company, GrowthExpectation is introduced as sixth variable. It represents the market-to-book ratio dividing market capitalization of a company by its book value of equity. (Cazavan-Jeny and Jeanjean, 2003; Fama and French, 1992; Markarian et al., 2008; Persson and Fuentes, 2011). It is expected that companies characterized by a high market-to-book ratio indicate a higher R&D intensity than companies having a low market value compared to their book value of equity (Cazavan-Jeny and Jeanjean, 2003). Additionally, it might be that companies which prefer to invest in growth opportunities (such as R&D) rather than in assets already in place show a higher market-to-book ratio since investors include those investments in their present value calculation leading to a higher share price. Due to these high investments company's debt might increase. This causes lenders to include higher (investment and financing) restrictions in their lending contracts limiting the firm's funding policy. These restrictions could refer to underinvestment issues or general accounting-based debt covenants. Obviously, managers want to avert the constraints using accounting discretion and thus tend to capitalize R&D to improve results and avoid showing underinvestment in their balance sheets. Thus, a positive relation between GrowthExpectation and RDCapitalization is hypothesized (Triki-Damak and Halioui,

2013; Myers, 1976; Shabou and Taktak, 2002).

Growth is introduced as seventh control variable, calculated as the percentage change in sales over one period. Growth is included as it is presumed that high growing companies engage in more R&D projects and thus invest more in R&D than companies with small growth rates (Cazavan-Jeny and Jeanjean, 2003; Persson and Fuentes, 2011).

The analysis additionally controls for years and industry effects by adding dummy variables to the regression. For each year and each industry, dummy variables are calculated while one of each category is excluded in the regression since there must be a reference category in case of categorial variables. For a detailed overview of the variables considered, including definition and formula, please refer to Appendix 1.

## 7. Data analysis and empirical results

### 7.1. Descriptive statistics

Table 2 shows the descriptive statistic on selected variables for the empirical analysis. Mean of RDCapitalization indicates that sample firms capitalize on average 1.61 % of their total assets. At the final ninetieth percentile, capitalization measures 5.24 % of total assets. Companies basically spend 6.03 % of their assets (assets are considered as before R&D capitalization as described in 6.2) on R&D. The most R&D intensive firms (last decile) even invest 15.97 % of their assets in R&D, while less R&D focused firms do not reach at least 1 %. Examining the relationship between RDCapitalization and RDTotals shows that companies capitalize on average 26.5 % of their R&D expenditures. The mean level of indebtedness reaches 65.69 %, indicating that selected companies are largely financed by debt rather than equity. The average (pre-managed) RoA shows that companies are profitable at 5.3 %, while the change in RoA clarifies that profitability of the sample firms decreases compared to the previous two fiscal years with a mean of -1.26 %. However, the first decile shows a positive RoA development at 2.21 %. The average market-to-book ratio is around 3.09 indicating that market capitalization of the analyzed companies is more than three times higher than their book value of equity. Thus, shareholders perceive these companies to be highly profitable. Even the lowest 10 % nearly show a ratio of one. Companies' growth in revenue reaches on average 6.85 % with a first decile of 23.07 %, indicating mostly positive growth rates. In the analysis Size, measured as the natural logarithm of total assets, indicates a mean of 9.37. Assets of the sample firms are on average 59 billion euros.

As the analysis is conducted across industries, it is appropriate to additionally perform a descriptive statistic by industry, presented in Table 3. Information technology and telecommunication services sectors show the highest R&D capitalization rates with 7.22 % and 4.10 % of their assets. These companies also record the highest R&D investments per assets indicating an average of 21.96 % and 11.57 %. The remaining industries vary between 5.81 % and 2.15 %. Companies of the materials sector and conglomerates reveal the



**Table 2:** Descriptive statistics on selected variables (own table)

|                       | Mean      | Std. Deviation | Minimum   | Maximum    | Percentiles |            |
|-----------------------|-----------|----------------|-----------|------------|-------------|------------|
|                       |           |                |           |            | 10          | 90         |
| RDCapitalization      | 0.016055  | 0.030586       | 0.000131  | 0.138229   | 0.000511    | 0.052395   |
| RDTotal               | 0.060349  | 0.093376       | 0.001134  | 0.752232   | 0.003046    | 0.159667   |
| Leverage              | 0.656868  | 0.154544       | 0.185897  | 0.929359   | 0.483224    | 0.844783   |
| ROA                   | 0.052995  | 0.046261       | -0.063809 | 0.153846   | -0.018890   | 0.110108   |
| $\Delta$ RoA          | -0.012644 | 0.042872       | -0.173077 | 0.046064   | -0.069908   | 0.022120   |
| GrowthExpectation     | 3.086864  | 3.281002       | 0.413424  | 27.923232  | 0.921325    | 5.049868   |
| Growth                | 0.068518  | 0.111889       | -0.244769 | 0.401038   | -0.056788   | 0.223071   |
| Size                  | 9.366865  | 2.330831       | 3.802208  | 12.909126  | 5.644880    | 12.039085  |
| Assets (EUR millions) | 59,010.97 | 88,264.17      | 51.90     | 409,732.00 | 306.88      | 171,077.20 |

lowest R&D capitalization rate with 0.07 % and 0.2 %. The most leveraged companies are those from the utilities and conglomerate sector with a debt to asset (before R&D) ratio of 78.89 % and 81.70 %. Companies in the health care sector are the least leveraged firms with debt of 54.72 % of their assets. All industries indicate an average positive profitability ratio before R&D capitalization, with the health care and materials sectors performing best at a pre-managed RoA of 8.6 %. Telecommunication services only reaches 0.32 % having the lowest profitability ratio before R&D capitalization. All industries, except health care, indicate a decreasing change in RoA compared to their previous two fiscal years with information technology companies having the lowest value (-8.2 %). This is consistent with the fact that these firms simultaneously show the highest R&D capitalization rate per assets. Their pre-managed RoA is relatively low compared to the operating profitability of previous years where R&D has already been capitalized. This leads to a highly negative change in RoA. Thus, these firms' performance largely depends on R&D capitalization.

## 7.2. Correlation analysis

In addition to the descriptive statistic, a correlation analysis (Table 4) is conducted to examine the (linear) relationships between the selected variables. As they are all scaled metrically, it is appropriate to perform a Bravais-Pearson correlation. It should be considered that Bravais-Pearson correlation analyzes associations at a univariate level. It does not incorporate cross-correlations between independent and control variables, which can change the results significantly. Additionally, the direction of the relationship cannot be determined. This calls for caution in interpreting the results. Furthermore, significant correlation does not necessarily indicate causality between variables (Markarian et al., 2008; Persson and Fuentes, 2011).

At first, the outcome statistically supports the hypothesis developed above since  $\Delta$ RoA and RDCapitalization show a strong and significant negative relationship. The same applies to pre-managed RoA being highly significant below the 0.01 level. This is visualized in Figure 2 showing a biaxial scatter diagram that plots the relationship between RoA and RDCapitalization as well as  $\Delta$ RoA and RDCapitalization.

Leverage and RDCapitalization show a positive correlation, supporting the assumption that managers also use R&D capitalization to avoid debt covenants violations. However, this relation is not significant. The strong positive relationship between RDTotal and RDCapitalization with high significance indicates that companies which invest more in R&D also tend to capitalize more as they have more resources available to finalize their projects (Figure 3).

It is interesting to observe that RDTotal and pre-managed RoA are significantly negatively correlated. This leads to the assumption that less profitable firms invest more in R&D than profitable ones. This might be due to the hope to increase their RoA when spending more on R&D. This guess stresses the importance of R&D investments from a firm's perspective. The same can be observed for the relation between  $\Delta$ RoA and RDTotal. The analysis also uncovers that small companies spend relatively more on R&D and tend to capitalize more than big companies since there is a negative significant relation between Size and RDTotal as well as Size and RDCapitalization. Finally, it could be observed that Growth and RDCapitalization are positively correlated. This demonstrates that companies being able to increase their sales compared to the previous year capitalize more of their R&D. This is in line with the correlation of RoA /  $\Delta$ RoA and RDCapitalization, since, ceteris paribus, a higher amount of revenue leads to a higher RoA. To decrease earnings in the current year, firms capitalize less to get closer to the prior years' earnings and thus smooth income.

Correlation by industry points out that the industries characterized by the highest capitalization rates, information technology and telecommunication services, also reveal the strongest, highly significant correlations between  $\Delta$ RoA and RDCapitalization. Simultaneously industries with low capitalization rates indicate low and insignificant relations between these two variables. The strongest relation between RDTotal and RDCapitalization can be found in telecommunication services and utilities industry with a correlation coefficient above 0.9. For further details on industry level a detailed Bravais-Pearson correlation among selected variables can be found in Appendix 2.

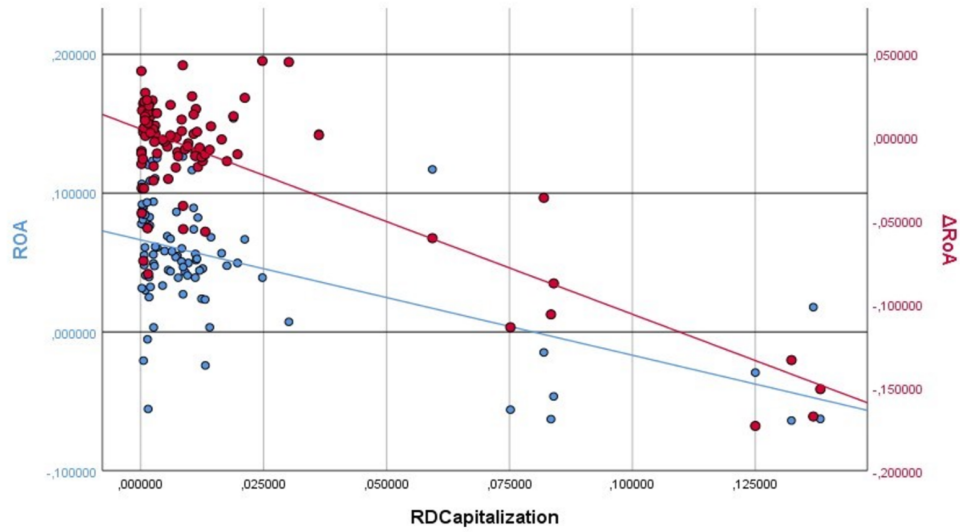
**Table 3:** Descriptive statistics on selected variables by industry (own table)

| Industry                   |                  | Mean      | Std. Devia-<br>tion | Minimum   | Maximum  |
|----------------------------|------------------|-----------|---------------------|-----------|----------|
| Industrials                | RDCapitalization | 0.009157  | 0.004916            | 0.003341  | 0.019698 |
|                            | RDTotal          | 0.029791  | 0.014085            | 0.010810  | 0.051334 |
|                            | Leverage         | 0.653007  | 0.104438            | 0.493048  | 0.780309 |
|                            | ROA              | 0.066361  | 0.025685            | 0.033399  | 0.126325 |
|                            | $\Delta$ RoA     | -0.002101 | 0.016233            | -0.024824 | 0.043442 |
| Health care                | RDCapitalization | 0.010442  | 0.016033            | 0.000278  | 0.059322 |
|                            | RDTotal          | 0.058173  | 0.053209            | 0.009190  | 0.162162 |
|                            | Leverage         | 0.547237  | 0.191795            | 0.185897  | 0.735515 |
|                            | ROA              | 0.086307  | 0.032081            | 0.027156  | 0.153846 |
|                            | $\Delta$ RoA     | 0.000621  | 0.022612            | -0.060360 | 0.020377 |
| Consumer discretionary     | RDCapitalization | 0.008643  | 0.004427            | 0.001499  | 0.014034 |
|                            | RDTotal          | 0.043423  | 0.019482            | 0.027697  | 0.080884 |
|                            | Leverage         | 0.725234  | 0.064679            | 0.594415  | 0.784248 |
|                            | ROA              | 0.052659  | 0.042292            | -0.024117 | 0.123249 |
|                            | $\Delta$ RoA     | -0.004819 | 0.017374            | -0.056443 | 0.022371 |
| Telecommunication services | RDCapitalization | 0.040983  | 0.043028            | 0.000704  | 0.083992 |
|                            | RDTotal          | 0.115746  | 0.122290            | 0.001134  | 0.238525 |
|                            | Leverage         | 0.618622  | 0.125519            | 0.488592  | 0.739036 |
|                            | ROA              | 0.003151  | 0.053632            | -0.062815 | 0.060898 |
|                            | $\Delta$ RoA     | -0.037811 | 0.056588            | -0.113842 | 0.021535 |
| Conglomerate               | RDCapitalization | 0.002017  | 0.000889            | 0.000567  | 0.002975 |
|                            | RDTotal          | 0.037886  | 0.028177            | 0.010108  | 0.098334 |
|                            | Leverage         | 0.817013  | 0.109941            | 0.701608  | 0.929359 |
|                            | ROA              | 0.035130  | 0.025970            | -0.020665 | 0.061358 |
|                            | $\Delta$ RoA     | -0.005757 | 0.031667            | -0.074000 | 0.026908 |
| Utilities                  | RDCapitalization | 0.011815  | 0.011909            | 0.001389  | 0.030133 |
|                            | RDTotal          | 0.021508  | 0.019598            | 0.002959  | 0.044223 |
|                            | Leverage         | 0.788916  | 0.094291            | 0.693869  | 0.896771 |
|                            | ROA              | 0.018339  | 0.039073            | -0.055501 | 0.066681 |
|                            | $\Delta$ RoA     | -0.000237 | 0.046842            | -0.081836 | 0.046064 |
| Materials                  | RDCapitalization | 0.000760  | 0.000926            | 0.000131  | 0.003303 |
|                            | RDTotal          | 0.024361  | 0.019280            | 0.001566  | 0.060960 |
|                            | Leverage         | 0.568811  | 0.051969            | 0.480923  | 0.642905 |
|                            | ROA              | 0.086136  | 0.022301            | 0.031566  | 0.125187 |
|                            | $\Delta$ RoA     | -0.004677 | 0.024695            | -0.045414 | 0.039978 |
| Information technology     | RDCapitalization | 0.072192  | 0.065243            | 0.008637  | 0.138229 |
|                            | RDTotal          | 0.219554  | 0.222554            | 0.094549  | 0.752232 |
|                            | Leverage         | 0.625258  | 0.234131            | 0.359281  | 0.917293 |
|                            | ROA              | 0.012970  | 0.057223            | -0.063809 | 0.073979 |
|                            | $\Delta$ RoA     | -0.081982 | 0.082921            | -0.173077 | 0.013906 |

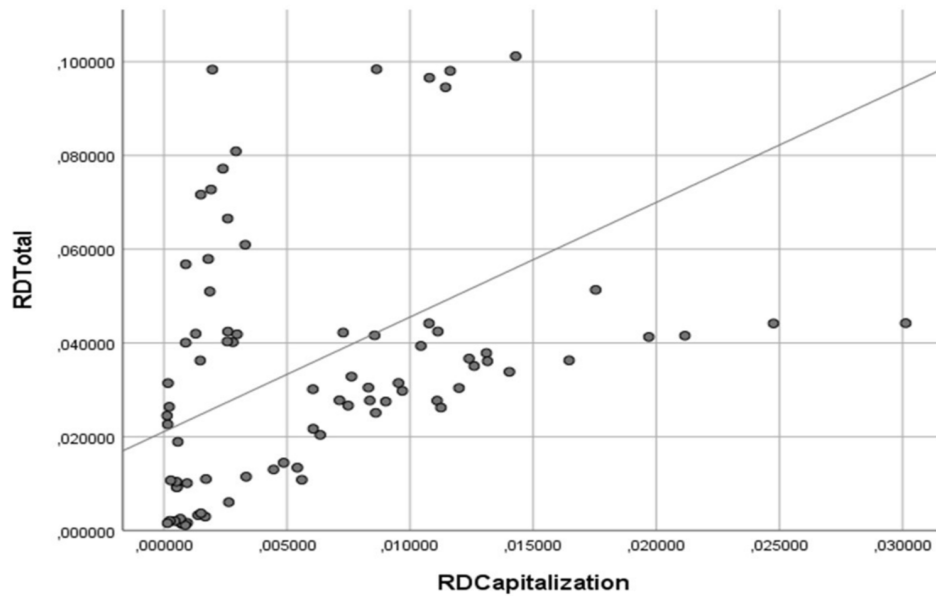
**Table 4:** Bravais-Pearson correlation among variables (own table);

\*, Correlation is significant at the 0.05 level (2-tailed).; \*\*, Correlation is significant at the 0.01 level (2-tailed).

|                   |                                        | Size<br>Capitalization | Lag<br>Capitalization | RD<br>Capitalizer | High              | RDTotal           | Leverage          | ROA              | $\Delta$ RoA   | Growth<br>Expectation |
|-------------------|----------------------------------------|------------------------|-----------------------|-------------------|-------------------|-------------------|-------------------|------------------|----------------|-----------------------|
| LagCapitalization | Pearson Correlation<br>Sig. (2-tailed) | -0.326**<br>0.002      |                       |                   |                   |                   |                   |                  |                |                       |
| RDCapitalization  | Pearson Correlation<br>Sig. (2-tailed) | -0.681**<br>0.000      | 0.517**<br>0.000      |                   |                   |                   |                   |                  |                |                       |
| HighCapitalizer   | Pearson Correlation<br>Sig. (2-tailed) | -0.388**<br>0.000      | 0.162<br>0.122        | 0.454**<br>0.000  |                   |                   |                   |                  |                |                       |
| RDTotal           | Pearson Correlation<br>Sig. (2-tailed) | -0.594**<br>0.000      | 0.455**<br>0.000      | 0.801**<br>0.000  | 0.326**<br>0.002  |                   |                   |                  |                |                       |
| Leverage          | Pearson Correlation<br>Sig. (2-tailed) | 0.401**<br>0.000       | -0.035<br>0.743       | 0.011<br>0.918    | -0.041<br>0.698   | -0.134<br>0.203   |                   |                  |                |                       |
| ROA               | Pearson Correlation<br>Sig. (2-tailed) | 0.084<br>0.425         | -0.196<br>0.062       | -0.549**<br>0.000 | -0.333**<br>0.001 | -0.260*<br>0.012  | -0.519**<br>0.000 |                  |                |                       |
| $\Delta$ RoA      | Pearson Correlation<br>Sig. (2-tailed) | 0.429**<br>0.000       | -0.481**<br>0.000     | -0.794**<br>0.000 | -0.277**<br>0.008 | -0.649**<br>0.000 | -0.122<br>0.245   | 0.671**<br>0.000 |                |                       |
| GrowthExpectation | Pearson Correlation<br>Sig. (2-tailed) | -0.359**<br>0.000      | 0.028<br>0.794        | 0.067<br>0.525    | -0.094<br>0.375   | 0.055<br>0.602    | -0.174<br>0.098   | 0.300**<br>0.004 | 0.137<br>0.194 |                       |
| Growth            | Pearson Correlation<br>Sig. (2-tailed) | -0.334**<br>0.001      | 0.030<br>0.779        | 0.228*<br>0.029   | 0.319**<br>0.002  | 0.133<br>0.208    | -0.114<br>0.281   | -0.001<br>0.989  | 0.170<br>0.106 | 0.234*<br>0.025       |



**Figure 2:** Scatter diagram between RoA,  $\Delta RoA$  and RDCapitalization (own figure)



**Figure 3:** Scatter diagram between RDTot and RDCapitalization (own figure)

### 7.3. Regression model

To test the hypothesis established, the statistical equation for the multiple linear regression has the following general form. The subscripted  $t$  represents different time periods, while  $i$  typifies each sample firm included in the analysis.

$$\begin{aligned} RDCapitalization_{it} = & b_{0it} + b_{1it}\Delta RoA_{it} + b_{3it}Leverage_{it} \\ & + b_{4it}RDTot_{it} + b_{5it}LagCapitalization_{it} \\ & + b_{6it}HighCapitalizer_{it} + b_{7it}Size_{it} + b_{8it}Growth_{it} \\ & + b_{9it}Growth_iExpectation_t + b_{10it}YearDummies_{it} \\ & + b_{11it}IndustryDummies_{it} + u_{it} \end{aligned} \quad (2)$$

According to the hypothesis developed in section 5, it is expected to observe a negative coefficient for  $b_{1it}$ . Table 5 presents the regression results.

Model 1 shows that  $\Delta RoA$  is significantly negative as predicted in hypothesis H. This indicates that companies tend to capitalize more if their current pre-managed profitability is lower than the average of the previous two fiscal years. Through R&D capitalization, managers avoid showing huge amounts of R&D investments on the income statement and thus improve their results and decrease the gap to prior years' earnings. In case of a higher profitability in the present year compared to the past, managers prefer to expense R&D since this helps them to lower their profitability and get closer to



**Table 5:** Model 1 - Multiple linear regression testing hypothesis H (own table);

The regression controls for years and industries; \*\*Significant at the 0.01 level / \*Significant at the 0.05 level; R = 0.956 / R Square = 0.914 / Adjusted R Square = 0.891 / Std. Error of the Estimate = 0.0101

|                   | Unstandardized Coefficients |            | Standardized Coefficients |        | Sig.  | 95% Confidence Interval for B |             |
|-------------------|-----------------------------|------------|---------------------------|--------|-------|-------------------------------|-------------|
|                   | B                           | Std. Error | Beta                      | t      |       | Lower Bound                   | Upper Bound |
| (Constant)        | 0.053                       | 0.014      |                           | 3.725  | 0.000 | 0.024                         | 0.081       |
| $\Delta$ RoA      | -0.158*                     | 0.065      | -0.222                    | -2.426 | 0.018 | -0.288                        | -0.028      |
| ROA               | -0.119*                     | 0.059      | -0.179                    | -1.998 | 0.049 | -0.237                        | 0.000       |
| Leverage          | 0.024*                      | 0.011      | 0.123                     | 2.161  | 0.034 | 0.002                         | 0.047       |
| RDTotal           | 0.075**                     | 0.021      | 0.228                     | 3.561  | 0.001 | 0.033                         | 0.116       |
| LagCapitalization | 0.058*                      | 0.023      | 0.107                     | 2.491  | 0.015 | 0.012                         | 0.105       |
| HighCapitalizer   | -0.002                      | 0.003      | -0.027                    | -0.468 | 0.641 | -0.008                        | 0.005       |
| Size              | -0.006**                    | 0.001      | -0.470                    | -5.742 | 0.000 | -0.008                        | -0.004      |
| Growth            | 0.016                       | 0.014      | 0.058                     | 1.121  | 0.266 | -0.012                        | 0.044       |
| GrowthExpectation | 0.000                       | 0.000      | 0.016                     | 0.341  | 0.734 | -0.001                        | 0.001       |

the reported earnings of previous years. This result clearly proves the assumption (developed in section five) that managers engage in accrual-based earnings management through R&D accounting discretion to smooth their income. The findings are consistent with prior studies conducted by Markarian et al. (2008), Triki-Damak and Halioui (2013) and Abrahams and Sidhu (1998) who also demonstrated that managers smooth income by exercising R&D accounting discretion. Garanina et al. (2016) also support the results in some parts as they find significant influence of  $\Delta$ ROA on RDCapitalization in Germany but not in Russia. Additionally, Persson and Fuentes (2011) confirm the outcomes at least in one of three periods observed. However, the study conducted by Guidara et al. (2014) was not able to prove that R&D accounting choice is a management tool to smooth income.

Moreover, Model 1 shows that RoA is negatively significant at a 0.05 level. This suggests that profitable firms (those with a high pre-managed RoA) capitalize less than companies with a lower profitability. The finding indirectly confirms the hypothesis H since it reveals that less profitable firms capitalize more of their R&D to compensate bad performance while highly profitable firms prefer expensing to offset results above average, i. a. to decrease political cost and fluctuation in their earnings (Aboody and Lev, 1998; Markarian et al., 2008; Persson and Fuentes, 2011). It also supports the assumption of Cazavan-Jeny and Jeanjean (2003) who argue that high profitable companies prefer to expense their R&D as they do not want to ruin the analysts' image concerning their earnings quality. The assumption that high profitability signals sufficient resources to finalize the R&D project and increases the probability to obtain future economic benefits has to be rejected (Markarian et al., 2008).

Leverage shows a positive and significant coefficient, indicating that the debt-covenant hypothesis can be confirmed as well. According to the results, managers do also use R&D accounting to avoid violating debt covenants (Aboody and Lev, 1998; Cazavan-Jeny and Jeanjean, 2003; Hamada, 1972). This is in line with studies conducted by Triki-Damak and

Halioui (2013). However, Markarian et al. (2008) have not found a significant association between Leverage and RDCapitalization. Garanina et al. (2016) could only support the debt covenant hypothesis for Russian but not for German companies.

Furthermore, RDTotal is significantly and positively related to RDCapitalization indicating that firms that invest more in R&D also capitalize larger parts of their R&D projects. Thus, it could be argued that the probability of R&D projects to fulfill the recognition criteria of IAS 38 increases if companies spend more on R&D. Nevertheless, it has to be taken into account that IAS 38 is subjected to accounting discretion (Garanina et al., 2016; Markarian et al., 2008). The results additionally show that companies do not tend to expense R&D due to the complexity to determine whether recognition criteria are met in case of higher R&D investments (Garanina et al., 2016; Markarian et al., 2008). Further, it is not confirmed that companies which spend less on R&D rather promote existing projects that increase the probability to fulfill recognition criteria.

As LagCapitalization reveals a positive coefficient, managers capitalize more in the current year if they did so in the previous period. It points out that companies attach importance to consistency and uniformity in their reporting behavior despite of underlying earnings management incentives (Markarian et al., 2008; Zicke, 2014).

HighCapitalizer is not significant, showing that the impact of capitalization on earnings does not determine R&D capitalization behavior (Markarian et al., 2008).

Model 1 proves that Size is negatively (significant) associated with RDCapitalization as well. This is in line with the assumption that large enterprises face higher market/political visibility, leading to stronger and more detailed examinations by several analysts. Thus, these companies cannot afford to heavily engage in earnings management (Garanina et al., 2016; Othman and Zéghal, 2006). Conversely, small companies that face less public attention can capitalize more and thus improve their results more easily. It further suggests

that huge companies mainly invest in basic research, maintenance and product upgrades that typically do not meet recognition criteria of IAS 38 (Cazavan-Jeny and Jeanjean, 2003; Persson and Fuentes, 2011). Another interpretation leads to the assumption that big enterprises expense larger parts of their R&D in order to avoid high political costs (Dufour and Zemzem, 2005). It is also possible that large companies undertake riskier projects than small firm do as managers' compensations are often based on stock options that correlate positively with the risk of underlying assets (Daves et al., 2000; Garanina et al., 2016).

Growth shows a positive but insignificant coefficient. Thus, it could not be assumed that growing companies that usually invest a lot in R&D tend to capitalize more due to increasing R&D resources available. (Cazavan-Jeny and Jeanjean, 2003; Persson and Fuentes, 2011).

Finally, GrowthExpectation is not significant as well. This shows that companies characterized by a high market-to-book ratio do not capitalize more, indicating that the stock market does not have an impact on capitalization behavior in this model. A detailed comparison of how the results of this study behave to prior literature can be found in Appendix 3.

To determine the quality of the regression model, adjusted  $R^2$  must be considered as the normal  $R^2$  automatically grows with increasing number of independent variables and therefore provides little informative value. Adjusted  $R^2$  indicates that 89.1 % of variance in RDCapitalization is explained by independent variables. Thus, the model appears to be a good estimation. The ANOVA table (Appendix 4) additionally shows that the regression creates significant results and can therefore be used as a valid model.

To examine whether some variables must be excluded due to collinearity, tolerance and variance inflation factor (VIF) are calculated. Table 6 points out that the regression does not face collinearity problems since tolerance exceeds the threshold of 0.1 and VIF is below 10 for all variables.

#### 7.4. Sensitivity tests

In the following section it will be examined whether changes in the regression model lead to fundamental changes in the results to determine the robustness of the analysis. Table 7 provides an overview of the tests that will be performed.

First, Model 2 is calculated replacing  $\Delta RoA$  by  $\Delta RoA2$ . In Model 1 it is assumed that managers as well as the market consider the change in operating profitability as the change in "base" earnings (Markarian et al., 2008). That is why  $\Delta RoA$  refers to the average of the previous two fiscal years. Nevertheless, it could also be expected that managers' smoothing target refers to the change related to the latest earnings reported. Therefore,  $\Delta RoA2$  is introduced as the change in operating profitability over one period (Markarian et al., 2008).

The results in table 8 show that  $\Delta RoA2$  is negative like in Model 1 but insignificant. Thus, the hypothesis H would be rejected in case of  $\Delta RoA2$ . Additionally, adjusted  $R^2$  decreases, indicating that the explanatory power in Model 1 is slightly higher. The outcomes suggest that the managers'

smoothing target refers to the change in "base" profits rather than to the change compared the previous year.

In Model 3 the dummy variable Manage is introduced to check for earnings thresholds. It equals to one if R&D capitalization helps managers to convert a negative change in operating profitability to a positive one. This might be relevant because managers who face difficulties to reach positive  $\Delta RoA$  could have an incentive to expense their R&D even though it is below the established targets, as achieving a positive  $\Delta RoA$  is impossible either way. Thus, Manage assesses whether R&D capitalization is influenced by the fact that the target of prior years' profitability is unattainable. It can be determined whether R&D capitalization is only driven by thresholds and benchmark beating or is also used for income smoothing purposes. Therefore Manage is expected to show a positive coefficient (Burgstahler and Dichev, 1997; Markarian et al., 2008; Zicke, 2014).

The results show that Manage is positively significant at a 0.01 level in both cases, indicating that the ability to offset a negative change in operating profitability through R&D capitalization has to be considered in managers' R&D capitalization behavior. Furthermore,  $\Delta RoA$  as well as  $\Delta RoA2$  are significant showing a negative coefficient. This proves that the income smoothing hypothesis is still confirmed even if benchmark-beating ability is taken into account. Adjusted  $R^2$  increases in both models by 1.8 % / 1 % showing that the explanatory power increases if Manage is added to the model. Therefore, it is appropriate to consider the additional control variable when examining R&D capitalization behavior. These findings are consistent with the study conducted by Markarian et al. (2008). In this model, income smoothing based on prior year's numbers turns out to be considerable which is previously rejected in Model 2. Even though explanatory power is higher in case of  $\Delta RoA$ , it could be suggested that managers focus on base profits as well as on profits of the prior year. As both income smoothing references seem to be decisive, this will also be considered in the following analysis.

Although most of the literature determines operating profitability as the income smoothing target as stated in section 3.2., there is still some disagreement about the primary smoothing target of a company. Therefore, a sensitivity analysis is used to evaluate whether managers might also use R&D capitalization in terms of other smoothing targets than operating profitability (Barnea et al., 1976; Makar et al., 2000; Markarian et al., 2008; Michelson et al., 1995; Moses, 1987). Basically, the decisive criterion to determine the variable to be smoothed is the management's assumption about how different stakeholders assess the smoothness of income. Obviously, this is not always clearly determined, leading to discussions and different perceptions about the smoothing object. This makes it difficult for researchers to operationalize the very general manner of the income smoothing hypothesis. Additionally, each company can pursue another smoothing target, which is of course not easy to identify (Imhoff, 1981; Ronen and Sadan, 1981). It might also be that managers act opportunistically and choose the smoothing target that maximizes their personal welfare. Based on

**Table 6:** Collinearity statistics (own table)

|                   | Tolerance | VIF   |
|-------------------|-----------|-------|
| $\Delta RoA$      | 0.143     | 6.974 |
| RoA               | 0.149     | 6.727 |
| Leverage          | 0.372     | 2.691 |
| RDTotal           | 0.293     | 3.410 |
| LagCapizalization | 0.651     | 1.537 |
| HighCapitalizer   | 0.372     | 2.689 |
| Size              | 0.179     | 5.601 |
| Growth            | 0.444     | 2.251 |
| GrowthExpectation | 0.543     | 1.840 |

**Table 7:** Overview of the sensitivity tests performed (own table);

\* Each model also tests the one year change and includes Manage as an additional variable. These variations can be found in the appendix.

| Models                                                          | Sensitivity test                                                                               |
|-----------------------------------------------------------------|------------------------------------------------------------------------------------------------|
| Model 2    Replacing $\Delta RoA$ by $\Delta RoA2$              | Considering a one year change in RoA instead of a two year change                              |
| Model 3    Introducing control variable Manage to Model 1 and 2 | Considering benchmark-beating ability                                                          |
| Model 4*    Replacing $\Delta RoA$ by $\Delta RoE$              | Considering the change in net income normalized by total equity as the income smoothing target |
| Model 5*    Replacing $\Delta RoA$ by $\Delta NetIncome/Assets$ | Considering the change in net income normalized by total assets as the income smoothing target |
| Model 6*    Replacing $\Delta RoA$ by $\Delta EPS$              | Considering the change in earnings per share as the income smoothing target                    |

**Table 8:** Model 2 - Replacing  $\Delta RoA$  by  $\Delta RoA2$  (own table);

The regression controls for years and industries; \*\*Significant at the 0.01 level / \*Significant at the 0.05 level; R = 0.954 / R Square = 0.911 / Adjusted R Square = 0.887 / Std. Error of the Estimate = 0.0103

|                   | Unstandardized Coefficients |            | Standardized Coefficients |        |       | 95% Confidence Interval for B |             |
|-------------------|-----------------------------|------------|---------------------------|--------|-------|-------------------------------|-------------|
|                   | B                           | Std. Error | Beta                      | t      | Sig.  | Lower Bound                   | Upper Bound |
| (Constant)        | 0.059                       | 0.014      |                           | 4.306  | 0.000 | 0.032                         | 0.087       |
| $\Delta RoA2$     | -0.101                      | 0.056      | -0.140                    | -1.814 | 0.074 | -0.211                        | 0.010       |
| ROA               | -0.164**                    | 0.053      | -0.247                    | -3.092 | 0.003 | -0.269                        | -0.058      |
| Leverage          | 0.022*                      | 0.011      | 0.113                     | 1.964  | 0.053 | 0.000                         | 0.045       |
| RDTotal           | 0.084**                     | 0.020      | 0.256                     | 4.111  | 0.000 | 0.043                         | 0.125       |
| LagCapizalization | 0.061**                     | 0.024      | 0.111                     | 2.513  | 0.014 | 0.013                         | 0.109       |
| HighCapitalizer   | -0.003                      | 0.003      | -0.045                    | -0.799 | 0.427 | -0.010                        | 0.004       |
| Size              | -0.006**                    | 0.001      | -0.479                    | -5.663 | 0.000 | -0.008                        | -0.004      |
| Growth            | 0.011                       | 0.014      | 0.038                     | 0.750  | 0.456 | -0.017                        | 0.038       |
| GrowthExpectation | 0.000                       | 0.000      | 0.013                     | 0.263  | 0.793 | -0.001                        | 0.001       |

**Table 9:** Model 3 - Introducing control variable Manage (own table);

The regression controls for years and industries; \*\*Significant at the 0.01 level / \*Significant at the 0.05 level; First table: R = 0.964 / R Square = 0.929 / Adjusted R Square = 0.909 / Std. Error of the Estimate = 0.0092; Second table: R = 0.959 / R Square = 0.920 / Adjusted R Square = 0.897 / Std. Error of the Estimate = 0.0098

|                   | Unstandardized Coefficients |            | Standardized Coefficients |        |       | 95% Confidence Interval for B |             |
|-------------------|-----------------------------|------------|---------------------------|--------|-------|-------------------------------|-------------|
|                   | B                           | Std. Error | Beta                      | t      | Sig.  | Lower Bound                   | Upper Bound |
| (Constant)        | 0.054                       | 0.013      |                           | 4.200  | 0.000 | 0.028                         | 0.080       |
| $\Delta RoA$      | -0.143*                     | 0.060      | -0.201                    | -2.400 | 0.019 | -0.262                        | -0.024      |
| ROA               | -0.098                      | 0.054      | -0.148                    | -1.800 | 0.076 | -0.207                        | 0.011       |
| Manage            | 0.012**                     | 0.003      | 0.165                     | 3.919  | 0.000 | 0.006                         | 0.018       |
| Leverage          | 0.023*                      | 0.010      | 0.117                     | 2.252  | 0.027 | 0.003                         | 0.044       |
| RDTotal           | 0.077**                     | 0.019      | 0.235                     | 4.030  | 0.000 | 0.039                         | 0.115       |
| LagCapitalization | 0.053*                      | 0.021      | 0.096                     | 2.454  | 0.017 | 0.010                         | 0.095       |
| HighCapitalizer   | -0.005                      | 0.003      | -0.088                    | -1.622 | 0.109 | -0.012                        | 0.001       |
| Size              | -0.006**                    | 0.001      | -0.483                    | -6.450 | 0.000 | -0.008                        | -0.004      |
| Growth            | 0.011                       | 0.013      | 0.039                     | 0.810  | 0.421 | -0.015                        | 0.037       |
| GrowthExpectation | 0.000                       | 0.000      | 0.027                     | 0.625  | 0.534 | -0.001                        | 0.001       |

|                   | Unstandardized Coefficients |            | Standardized Coefficients |        |       | 95% Confidence Interval for B |             |
|-------------------|-----------------------------|------------|---------------------------|--------|-------|-------------------------------|-------------|
|                   | B                           | Std. Error | Beta                      | t      | Sig.  | Lower Bound                   | Upper Bound |
| (Constant)        | 0.053                       | 0.013      |                           | 3.980  | 0.000 | 0.026                         | 0.079       |
| $\Delta RoA2$     | -0.125*                     | 0.054      | -0.174                    | -2.331 | 0.023 | -0.232                        | -0.018      |
| RoA               | -0.131**                    | 0.052      | -0.198                    | -2.527 | 0.014 | -0.234                        | -0.028      |
| Manage2           | 0.009**                     | 0.003      | 0.120                     | 2.842  | 0.006 | 0.003                         | 0.015       |
| Leverage          | 0.024*                      | 0.011      | 0.122                     | 2.221  | 0.030 | 0.002                         | 0.046       |
| RDTotal           | 0.073**                     | 0.020      | 0.224                     | 3.690  | 0.000 | 0.034                         | 0.113       |
| LagCapitalization | 0.056*                      | 0.023      | 0.102                     | 2.412  | 0.018 | 0.010                         | 0.102       |
| HighCapitalizer   | -0.003                      | 0.003      | -0.057                    | -1.049 | 0.298 | -0.010                        | 0.003       |
| Size              | -0.006**                    | 0.001      | -0.462                    | -5.715 | 0.000 | -0.008                        | -0.004      |
| Growth            | 0.006                       | 0.013      | 0.020                     | 0.408  | 0.684 | -0.021                        | 0.032       |
| GrowthExpectation | 0.000                       | 0.000      | 0.028                     | 0.609  | 0.544 | -0.001                        | 0.001       |

that assumption, managers who are compensated dependent on net income are less interested in smoothing net income than managers getting compensated by the market value of the company. It has to be considered that all these factors again depend on how much influence these managers have on income smoothing (Imhoff, 1981). As it becomes clear that the income smoothing target is not easy to determine, the sensitivity analysis tests different smoothing objectives to examine which variable is most likely to be smoothed without specifying that this is the only target used across all firms. This extends prior literature on income smoothing through R&D accounting since it solely focused on operating profitability.

Model 4 replaces  $\Delta RoA$  by  $\Delta RoE$  (Return on equity), which is calculated as the difference between current pre-managed net income before R&D capitalization and the average net income the previous two fiscal years normalized by

current year's equity.

$$\Delta RoE_t = \frac{(\text{NetIncome}_t - \text{R\&D capitalized}_t) - \frac{\text{NetIncome}_{t-1} + \text{NetIncome}_{t-2}}{2}}{\text{Equity}_t} \quad (3)$$

Net income is chosen because prior research on income smoothing suggests that managers who are concerned with long term performance might prefer net income as the appropriate target (Ronen and Sadan, 1981). It must be considered that tax effects are neglected in this model since pre-managed net income is only calculated by subtracting R&D capitalization from bottom line profit without taking the tax effects incorporated into account (R&D reduces earnings before tax and thus tax payments). Table 10 shows the results

after replacing  $\Delta\text{RoA}$  by  $\Delta\text{RoE}$ .

The regression reveals that  $\Delta\text{RoE}$  is also highly negatively significant and thus an essential factor that has to be considered when analyzing R&D capitalization behavior. The model indicates that managers also focus on income smoothing in terms of bottom line profit. In consequence, operating profitability cannot be assumed to be the only smoothing target of a company. Adjusted  $R^2$  decreases by nearly two percentage points compared to Model 1 indicating that  $\Delta\text{RoA}$  is somewhat more important than  $\Delta\text{RoE}$  in terms of income smoothing. RoE, Leverage, RDTot, LagCapitalization and Size remain significant with the same coefficient signs as in Model 1. Adding Manage (related to  $\Delta\text{RoE}$ ) to the Model as performed in Model 3, shows that adjusted  $R^2$  increases with Manage being positively significant (Appendix 5). This is consistent with findings in the previous model. In conclusion, there are no fundamental changes if  $\Delta\text{RoA}$  is replaced by  $\Delta\text{RoE}$ .

In a second step,  $\Delta\text{RoE}$  is substituted by  $\Delta\text{RoE2}$  which measures the change in RoE over one period to face the issue discussed in Model 2.  $\Delta\text{RoE2}$  also shows a significant negative coefficient indicating that managers are not only concerned with base RoE but also with the change to the previous year. Adjusted  $R^2$  increases marginally, showing that  $\Delta\text{RoE2}$  explains R&D capitalization variability slightly better than  $\Delta\text{RoE}$ . As in the previous models, including Manage (related to  $\Delta\text{RoE2}$ ) increases explanatory power once more by nearly 1 %. The respective regressions can be found in Appendix 6.

Model 5 uses nearly the same variable definition as Model 4. The only difference is that the change in net income is normalized by total assets before R&D capitalization.

$$\Delta\text{NetIncome}/\text{Assets}_t = \frac{(\text{NetIncome}_t - \text{R\&D capitalized}_t) - \frac{\text{NetIncome}_{t-1} + \text{NetIncome}_{t-2}}{2}}{\text{EquityTotal assets}_t - \text{R\&D capitalized}_t} \quad (4)$$

Table 11 shows the results after replacing  $\Delta\text{RoA}$  by  $\Delta\text{NetIncome}/\text{Assets}$ .

As expected, the slight adjustment in Model 5 does not change the results fundamentally.  $\Delta\text{NetIncome}/\text{Assets}$  is still negatively significant. However, adjusted  $R^2$  increases by more than 2.4 percentage points up to 0.915 compared to Model 1. This indicates a very good explanatory power of the variables included. Consequently, the outcomes suggest that net income and its change normalized by total assets before R&D capitalization might even be more important concerning management's smoothing target than operating profitability. Otherwise, the findings equal those in Model 1 and Model 4 in terms of significance and the variables' coefficient signs except Leverage which is no longer significant. Also, in this model Manage (related to  $\Delta\text{NetIncome}/\text{Assets}$ ) is positively significant with an increasing adjusted  $R^2$  (Appendix 7). Regressing for  $\Delta\text{NetIncome}/\text{Assets2}$ , which again

looks at one-year changes shows a negative and significant coefficient with the same adjusted  $R^2$  as in Model 5. Thus,  $\Delta\text{NetIncome}/\text{Assets}$  can be considered equally important to managers as  $\Delta\text{NetIncome}/\text{Assets2}$ . Manage (related to  $\Delta\text{NetIncome}/\text{Assets2}$ ) additionally increases explanatory power by 1 %. The regressions for the second analysis can be found in Appendix 8.

The last model uses  $\Delta\text{EPS}$  as the smoothing variable. It is calculated by dividing net income before R&D capitalization by the total number of shares. This measure is introduced since some prior income smoothing literature suggest to use EPS as an income smoothing object because of its importance in annual reports and several analysis (White, 1970).

$$\Delta\text{EPS}_t = \frac{(\text{NetIncome}_t - \text{R\&D capitalized}_t)}{\text{Total numbers of shares}_t} - \frac{\frac{(\text{NetIncome}_{t-1})}{\text{Total numbers of shares}_{t-1}} + \frac{(\text{NetIncome}_{t-2})}{\text{Total numbers of shares}_{t-2}}}{2} \quad (5)$$

Table 12 illustrates the results after replacing  $\Delta\text{RoA}$  by  $\Delta\text{EPS}$ .

The outcomes suggest that  $\Delta\text{EPS}$  is also negatively significant as in prior models. Therefore, it can be assumed that  $\Delta\text{EPS}$  is integrated in managers' smoothing targets too. This model indicates the lowest adjusted  $R^2$  with 0.856. Even though explanatory power is still high, it decreased by 3.5 % compared to Model 1. Like in Model 4 and Model 1, Leverage becomes significant when taking ESP into account, supporting the debt covenant hypothesis. Pre-managed ESP however is not significant indicating that companies characterized by a low ESP do not capitalize more to improve their EPS. It seems that companies are only interested in smooth EPS but are rather less concerned with the absolute magnitude. Further, Manage (related to  $\Delta\text{EPS}$ ) increases adjusted  $R^2$  by nearly 2 % (Appendix 9). As in the previous models,  $\Delta\text{EPS}$  is replaced by  $\Delta\text{EPS2}$  to assess whether base changes or one-year changes are perceived as more important.  $\Delta\text{EPS2}$  is significantly negative associated to R&D capitalization with a slightly decreasing adjusted  $R^2$  compared to Model 6. Manage (related to  $\Delta\text{EPS2}$ ) leads to improved explanatory power of 3.3 %. The analysis can be found in Appendix 10.

Table 13 summarizes the results of the sensitivity analysis. The outcomes suggest that managers do not only use operating profitability as the income smoothing target as assumed by prior R&D capitalization studies.

Net income as well as EPS are also considered by managers in terms of income smoothing. The findings extend prior literature on that field since no study investigated R&D capitalization based on net income or EPS smoothing even though these variables are highly significant and therefore have to be taken into account. Additionally, it can be proofed that companies do not only focus on base numbers but also consider the change to the prior year. Adding Manage to the models consistently increases explanatory power indicating that benchmark-beating ability should always be considered.



**Table 10:** Model 4 - Replacing  $\Delta RoA$  by  $\Delta RoE$  (own table)

The regression controls for years and industries; \*\*Significant at the 0.01 level / \*Significant at the 0.05 level ; R = 0.949 / R Square = 0.901/ Adjusted R Square = 0.874 / Std. Error of the Estimate = 0.0108

|                   | Unstandardized Coefficients |            | Standardized Coefficients |        | Sig.  | 95% Confidence Interval for B |             |
|-------------------|-----------------------------|------------|---------------------------|--------|-------|-------------------------------|-------------|
|                   | B                           | Std. Error | Beta                      | t      |       | Lower Bound                   | Upper Bound |
| (Constant)        | 0.048                       | 0.012      |                           | 4.122  | 0.000 | 0.025                         | 0.071       |
| $\Delta RoE$      | -0.024**                    | 0.008      | -0.164                    | -3.139 | 0.002 | -0.040                        | -0.009      |
| RoE               | -0.025*                     | 0.011      | -0.130                    | -2.342 | 0.022 | -0.046                        | -0.004      |
| Leverage          | 0.040**                     | 0.011      | 0.201                     | 3.598  | 0.001 | 0.018                         | 0.062       |
| RDTotal           | 0.073**                     | 0.021      | 0.224                     | 3.445  | 0.001 | 0.031                         | 0.116       |
| LagCapitalization | 0.069**                     | 0.024      | 0.126                     | 2.829  | 0.006 | 0.020                         | 0.117       |
| HighCapitalizer   | 0.002                       | 0.003      | 0.029                     | 0.507  | 0.614 | -0.005                        | 0.009       |
| Size              | -0.007**                    | 0.001      | -0.564                    | -6.690 | 0.000 | -0.010                        | -0.005      |
| Growth            | 0.002                       | 0.014      | 0.007                     | 0.127  | 0.899 | -0.026                        | 0.030       |
| GrowthExpectation | 0.000                       | 0.000      | -0.038                    | -0.766 | 0.446 | -0.001                        | 0.001       |

**Table 11:** Model 5 - Replacing  $\Delta RoA$  by  $\Delta NetIncome/Assets$  (own table)

The regression controls for years and industries; \*\*Significant at the 0.01 level / \*Significant at the 0.05 level ; R = 0.966 / R Square = 0.933/ Adjusted R Square = 0.915 / Std. Error of the Estimate = 0.0089

|                           | Unstandardized Coefficients |            | Standardized Coefficients |        | Sig.  | 95% Confidence Interval for B |             |
|---------------------------|-----------------------------|------------|---------------------------|--------|-------|-------------------------------|-------------|
|                           | B                           | Std. Error | Beta                      | t      |       | Lower Bound                   | Upper Bound |
| (Constant)                | 0.059                       | 0.011      |                           | 5.251  | 0.000 | 0.037                         | 0.082       |
| $\Delta NetIncome/Assets$ | -0.154**                    | 0.051      | -0.220                    | -3.036 | 0.003 | -0.254                        | -0.053      |
| NetIncome/Assets          | -0.201**                    | 0.059      | -0.284                    | -3.386 | 0.001 | -0.319                        | -0.083      |
| Leverage                  | 0.005                       | 0.013      | 0.024                     | 0.367  | 0.715 | -0.021                        | 0.031       |
| RDTotal                   | 0.059**                     | 0.019      | 0.181                     | 3.175  | 0.002 | 0.022                         | 0.097       |
| LagCapitalization         | 0.058**                     | 0.020      | 0.107                     | 2.889  | 0.005 | 0.018                         | 0.099       |
| HighCapitalizer           | 0.000                       | 0.003      | 0.003                     | 0.069  | 0.945 | -0.005                        | 0.006       |
| Size                      | -0.006**                    | 0.001      | -0.430                    | -5.997 | 0.000 | -0.008                        | -0.004      |
| Growth                    | 0.007                       | 0.012      | 0.027                     | 0.638  | 0.525 | -0.016                        | 0.031       |
| GrowthExpectation         | 0.000                       | 0.000      | -0.015                    | -0.357 | 0.722 | -0.001                        | 0.001       |

**Table 12:** Model 6 - Replacing  $\Delta RoA$  by  $\Delta EPS$  (own table)

The regression controls for years and industries; \*\*Significant at the 0.01 level / \*Significant at the 0.05 level ; R = 0.942 / R Square = 0.887/ Adjusted R Square = 0.856 / Std. Error of the Estimate = 0.0117

|                   | Unstandardized Coefficients |            | Standardized Coefficients |        | Sig.  | 95% Confidence Interval for B |             |
|-------------------|-----------------------------|------------|---------------------------|--------|-------|-------------------------------|-------------|
|                   | B                           | Std. Error | Beta                      | t      |       | Lower Bound                   | Upper Bound |
| (Constant)        | 0.066                       | 0.016      |                           | 4.019  | 0.000 | 0.033                         | 0.098       |
| $\Delta EPS$      | -0.001**                    | 0.000      | -0.119                    | -2.609 | 0.011 | -0.001                        | 0.000       |
| EPS               | 0.001                       | 0.000      | 0.092                     | 1.415  | 0.161 | 0.000                         | 0.002       |
| Leverage          | 0.056**                     | 0.011      | 0.284                     | 5.249  | 0.000 | 0.035                         | 0.078       |
| RDTotal           | 0.076**                     | 0.023      | 0.231                     | 3.359  | 0.001 | 0.031                         | 0.121       |
| LagCapitalization | 0.081**                     | 0.026      | 0.148                     | 3.067  | 0.003 | 0.028                         | 0.133       |
| HighCapitalizer   | 0.000                       | 0.004      | -0.004                    | -0.059 | 0.953 | -0.008                        | 0.007       |
| Size              | -0.009**                    | 0.001      | -0.693                    | -7.846 | 0.000 | -0.011                        | -0.007      |
| Growth            | -0.017                      | 0.015      | -0.063                    | -1.190 | 0.238 | -0.046                        | 0.012       |
| GrowthExpectation | 0.000                       | 0.000      | -0.052                    | -0.976 | 0.333 | -0.001                        | 0.001       |

**Table 13:** Summary of results of sensitivity analysis (own table)

| Variables                                | Result                   | Significance level | Adjusted R Square |
|------------------------------------------|--------------------------|--------------------|-------------------|
| $\Delta RoA$                             | significantly negative   | 0.018              | 0.891             |
| $\Delta RoA$ (inc. Manage)               | significantly negative   | 0.019              | 0.909             |
| $\Delta RoA2$                            | insignificantly negative | 0.074              | 0.887             |
| $\Delta RoA2$ (inc. Manage)              | significantly negative   | 0.023              | 0.897             |
| $\Delta RoE$                             | significantly negative   | 0.002              | 0.874             |
| $\Delta RoE$ (inc. Manage)               | significantly negative   | 0.007              | 0.881             |
| $\Delta RoE2$                            | significantly negative   | 0.000              | 0.877             |
| $\Delta RoE2$ (inc. Manage)              | significantly negative   | 0.001              | 0.885             |
| $\Delta NetIncome/Assets$                | significantly negative   | 0.003              | 0.915             |
| $\Delta NetIncome/Assets$ (inc. Manage)  | significantly negative   | 0.008              | 0.920             |
| $\Delta NetIncome/Assets2$               | significantly negative   | 0.003              | 0.915             |
| $\Delta NetIncome/Assets2$ (inc. Manage) | significantly negative   | 0.002              | 0.926             |
| $\Delta EPS$                             | significantly negative   | 0.011              | 0.856             |
| $\Delta EPS$ (inc. Manage)               | significantly negative   | 0.001              | 0.873             |
| $\Delta EPS2$                            | significantly negative   | 0.033              | 0.852             |
| $\Delta EPS2$ (inc. Manage)              | significantly negative   | 0.018              | 0.885             |

## 8. Limitations and further research potential

Even though the analysis proofed that managers engage in income smoothing through R&D accounting discretion, there are some limitations of the research that have to be regarded.

First, the thesis only focuses on accrual-based earnings management to smooth income while neglecting real earnings management, even though prior studies indicate that the latter does influence R&D expenditures (Cazavan-Jeny et al., 2011; Markarian et al., 2008; Osmia and Young, 2009; Seybert, 2010; Tahinakis, 2014). If it is assumed that managers prefer to regulate their income by adopting R&D expenditures rather than change R&D capitalization ratio, this would impair the results leading to a lower association between the variables expatiated and R&D capitalization (Markarian et al., 2008). Since IFRS's aim is to provide decision useful information, primary to investors, it supports investor protection. Hence, companies might choose to favor real earnings management since this is less observable than accrual-based earnings management (Enomoto et al., 2015).

Second, R&D related data has to be collected by hand. This is very time consuming and therefore leads to a relatively low sample size. Additionally, data collection was partly restricted by disclosing issues, since not all companies publish their R&D numbers concretely and therefore have to be excluded from the sample. Beside of the very elaborate preselection of companies that are suitable for the analysis over the examined period, the elimination of the companies which do not disclose R&D numbers appropriately might cause a potential bias in the results since these companies could also be engaged in earnings management (Persson and Fuentes, 2011).

Additionally, the analysis only focuses on German listed companies which limits the validity to only one country. Since research results on that topic are quite diverse and

vary by the country examined, further research could explore how the findings change if the analysis is conducted including multiple countries. The prevailing accounting and financial conditions as well as the level of control and regulation in a country might influence managers' accounting behavior and targets. Especially developing countries with rising R&D intensity and globalization effects could be interesting to observe in the future. Another possible research topic could be based on a comparison between two countries that differ in how to account for R&D. If a country with R&D capitalization options is opposed to a country that demands expensing of all R&D, the results determine the importance and role of accruals (Garanina et al., 2016; Triki-Damak and Halioui, 2013).

Further, there are some mediating and moderating effects that were not considered and might alter the results. The thesis does not incorporate the economic situation in the years examined even though it is possible that the accounting behavior of companies changes based on economic development (Garanina et al., 2016; Persson and Fuentes, 2011). In addition, it would be worth to consider Wang et al. (2017) suggestion that family-owned companies tend to show a higher quality of earnings than other stock companies. Consequently, it could be assumed that family owned business are less engaged in earnings management. Other factors that are suggested to influence the results are board size and its independence as these are considered as influencing factors in terms of monitoring effectiveness. The smaller the board and the higher its independence the better the associated monitoring quality. It could be assumed that a better monitoring quality allows for less earnings management through accounting discretion (Klein, 2002; Makar et al., 2000; Vafeas, 2000). Another aspect of interest might be the compensation of executives since the bonus plan hypothesis suggests that managers conduct earnings management to increase their personal wealth if their compensation is based

on accounting numbers (Garanina et al., 2016; Triki-Damak and Halioui, 2013; Makar et al., 2000). The analysis also neglects the role of auditors. The quality of audit might be an additional variable of interest since a detailed and good external audit allows for less earnings management possibilities. However, it is difficult to quantify a high or low audit quality (Triki-Damak and Halioui, 2013; Othman and Zéghal, 2006). In fact, these factors are not considered because they are mostly not disclosed in the annual reports of the companies examined. Despite the fact that all these potentially influential factors are not included, the explanatory power of the present analysis is very high (always above 80 %) indicating that most of the significant variables that determine R&D capitalization behavior have already been identified.

## 9. Conclusion

The thesis examines whether managers of German listed companies use the flexibility to capitalize R&D expenditures given by IAS 38 to engage in income smoothing in terms of accrual-based earnings management. It is argued that there is a negative relationship between the companies' change in operating profitability and R&D capitalization.

To investigate the association between R&D capitalization and income smoothing, a multiple linear regression analysis is performed including relevant control variables. The outcomes support the established hypothesis, since  $\Delta\text{RoA}$  shows a negative and significant coefficient. This proves that managers do conduct earnings management through R&D capitalization to smooth their incomes. Further, a set of sensitivity analysis is conducted. The findings indicate that the hypothesis is still confirmed even if benchmark-beating ability is considered. Explanatory power even increases when adding the new variable. Additionally, the analysis proves that operating profitability could not be presumed to be the only income smoothing target since  $\Delta\text{RoE}$ ,  $\Delta\text{NetIncome}/\text{Assets}$  and  $\Delta\text{EPS}$  are significant as well. This extends prior literature as they only focus on operating profitability. Further, the results suggest that income smoothing could refer to base profits as well as to profits of the prior year.

The analysis contributes to the convergence project between the United States Generally Accepted Accounting Principles (US-GAAP) and IFRS, signed in 2002 (Cheney, 2017). It supports the view of FASB arguing that R&D capitalization opportunities pave the way for earnings management and thus impair objectivity of financial statements (Cazavan-Jeny et al., 2011; Healy et al., 2002; Lev and Sougiannis, 1996). However, while full expensing of R&D prevents managers from accrual-based earnings management, it could not be ensured that companies will not conduct real earnings management by adjusting their total amount of R&D investments made, according to current performance. If companies reduce their R&D spending in times of low profitability this is counterproductive since new technologies might help them to improve their business. Hence, US-GAAP might prevent companies from investing in R&D even though these investments are vital for today's companies to survive. In con-

sequence, FASB and ISAB have to find a way of accounting practice that does not impair objectivity, maintain reliability and support R&D investments since a knowledge-based society forces companies to come up with new developments and improvements. To reach that, standard setters, politics and the companies itself have to work together to cope with the technology-caused structural change that is currently observed in the economy.

## References

- Aboudy, D. and Lev, B. The value relevance of intangibles: The case of software capitalization. *Journal of Accounting Research*, 36:161–191, 1998.
- Abrahams, T. and Sidhu, B. K. The role of R&D capitalisations in firm valuation and performance measurement. *Australian Journal of Management*, 23(2):169–183, 1998.
- Barnea, A., Ronen, J., and Sadan, S. Classificatory smoothing of income with extraordinary items. *The Accounting Review*, 51(1):110–122, 1976.
- Beidleman, C. R. Income smoothing: The role of management. *The Accounting Review*, 48(4):653–667, 1973.
- Boone, A. L., Field, L. C., Karpoff, J. M., and Raheja, C. G. The determinants of corporate board size and composition: An empirical analysis. *Journal of Financial Economics*, 85(1):66–101, 2007.
- Boulton, R., Libert, B., and Samek, S. M. Cracking the value code: How successful businesses are creating wealth in the new economy. New York: HarperBusiness, 2000.
- Brayshaw, R. and Eldin, A. E. The smoothing hypothesis and the role of exchange differences. *Journal of Business Finance & Accounting*, 16(5): 621–633, 1989.
- Burgstahler, D. and Dichev, I. Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics*, 24(1):99–126, 1997.
- Cazavan-Jeny, A. and Jeanjean, T. Value Relevance of R&D Reporting: A Signaling Interpretation, 2003.
- Cazavan-Jeny, A., Jeanjean, T., and Joos, P. Accounting choice and future performance: The case of R&D accounting in France. *Journal of Accounting and Public Policy*, 30(2):145–165, 2011.
- Chan, H. W., Faff, R. W., Gharghori, P., and Ho, Y. K. The relation between R&D intensity and future market returns: does expensing versus capitalization matter? *Review of Quantitative Finance and Accounting*, 29(1): 25–51, 2007.
- Cheney, T. GAAP and IFRS Convergence Uncovered - Big 4 Exploration Series, 2017. URL <https://www.bna.com/gaap-ifrs-convergence-b73014463664/>.
- Daves, P. R., Ehrhardt, M. C., Kuhlemeyer, G. A., and Kunkel, R. A. Increases in the systematic risk of large firms. *American Business Review*, 18(2): 62–74, 2000.
- Dufour, D. and Zemzem, A. Activation des frais de R&D: le cas des entreprises françaises cotées sur le Nou-veau Marché, 2005. URL <https://halshs.archives-ouvertes.fr/halshs-00581187/document>.
- Duke, J. C. and Hunt III, H. G. An empirical examination of debt covenant restrictions and accounting-related debt proxies. *Journal of Accounting and Economics*, 12(1-3):45–63, 1990.
- Enomoto, M., Kimura, F., and Yamaguchi, T. Accrual-based and real earnings management: An international comparison for investor protection. *Journal of Contemporary Accounting & Economics*, 11(3):183–198, 2015.
- Fama, E. F. and French, K. R. The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465, 1992.
- Garanina, T., Nikulin, E., and Frangulants, O. Earnings management and R&D costs capitalization: evidence from Russian and German markets. *Investment Management and Financial Innovations*, 13(1):77–85, 2016.
- Guidara, R., Boujelbene, Y., et al. R&D-based earnings management and accounting performance motivation. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, 4(2):81–93, 2014.
- Hadjiloucas, T. The increasing importance of brand and intangibles in industry, 2014. URL [http://www.wipo.int/edocs/mdocs/sme/en/wipo\\_mes\\_lis\\_14/wipo\\_mes\\_lis\\_14\\_g\\_hadjiloucas.pdf](http://www.wipo.int/edocs/mdocs/sme/en/wipo_mes_lis_14/wipo_mes_lis_14_g_hadjiloucas.pdf).
- Hamada, R. S. The effect of the firm's capital structure on the systematic risk of common stocks. *The Journal of Finance*, 27(2):435–452, 1972.
- Healy, P. M. and Wahlen, J. M. A review of the earnings management literature and its implications for standard setting. *Accounting Horizons*, 13(4):365–383, 1999.
- Healy, P. M., Myers, S. C., and Howe, C. D. R&D accounting and the tradeoff between relevance and objectivity. *Journal of Accounting Research*, 40(3): 677–710, 2002.
- Hepworth, S. R. Smoothing periodic income. *The Accounting Review*, 28(1): 32–39, 1953.
- Hernández, H., Grassano, N., Tübke, A., Potters, L., Amoroso, S., Dosso, M., and Vezzani, A. The 2017 EU Industrial R&D Investment Scoreboard, 2017.
- Hoffmann, C. Die Industrie stirbt. *Süddeutsche.de*, 1–6., 2017.
- Imhoff, E. A. Income Smoothing: An Analysis of Critical Issues. *Quarterly Review of Economics and Business* (Autumn), 21(3), 23–42., 1981.
- Industrial Research Institute. Research & Development - 2017 Global R&D Funding Forecast. *R&D Magazine*, 1–36., 2017.
- Jones, J. J. Earnings management during import relief investigations. *Journal of Accounting Research*, 29(2):193–228, 1991.
- Klein, A. Audit committee, board of director characteristics, and earnings management. *Journal of Accounting and Economics*, 33(3):375–400, 2002.
- Koalitionsvertrag. Ein neuer Aufbruch für Europa Eine neue Dynamik für Deutschland Ein neuer Zusammenhalt für unser Land Koalitionsvertrag zwischen CDU, CSU und SPD., 2018. URL [https://www.handelsblatt.com/downloads/20936422/4/koalitionsvertrag\\_final.pdf](https://www.handelsblatt.com/downloads/20936422/4/koalitionsvertrag_final.pdf).
- Koch, B. S. Income smoothing: An experiment. *The Accounting Review*, 56(3):574, 1981.
- Leuz, C., Nanda, D., and Wysocki, P. D. Earnings management and investor protection: an international comparison. *Journal of Financial Economics*, 69(3):505–527, 2003.
- Lev, B. and Sougiannis, T. The capitalization, amortization, and value-relevance of R&D., 1996.
- Liberty, S. E. and Zimmerman, J. L. Labor union contract negotiations and accounting choices. *Accounting Review*, pages 692–712, 1986.
- Makar, S. D., Alam, P., and Pearson, M. A. Earnings Management, 2000. URL <http://www.fraud-magazine.com/article.aspx?id=4294968448>.
- Markarian, G., Pozza, L., and Prencipe, A. Capitalization of R&D costs and earnings management: Evidence from Italian listed companies. *The International Journal of Accounting*, 43(3):246–267, 2008.
- Marx, K. Grundrisse: Outlines of the Critique of Political Economy., 1857. URL [https://www.marxists.org/archive/marx/works/download/Marx\\_Grundrisse.pdf](https://www.marxists.org/archive/marx/works/download/Marx_Grundrisse.pdf).
- Michelson, S. E., Jordan-Wagner, J., and Wootton, C. W. A market based analysis of income smoothing. *Journal of Business Finance & Accounting*, 22(8):1179–1193, 1995.
- Miller, G. S. and Skinner, D. J. Determinants of the valuation allowance for deferred tax assets under sfas no. 109. *Accounting Review*, pages 213–233, 1998.
- Moses, O. D. Income smoothing and incentives: Empirical tests using accounting changes. *Accounting Review*, pages 358–377, 1987.
- Myers, S. C. Determinants of corporate borrowing. *Journal of Financial Economics*, 1–47., 1976.
- Nelson, M. W., Elliott, J. A., and Tarpley, R. L. How earnings are managed Examples from auditors. *Accounting Horizons* (Supplement), 17–35., 2003.
- Osma, B. G. and Young, S. R&D expenditure and earnings targets. *European Accounting Review*, 18(1):7–32, 2009.
- Oswald, D. R. and Zarowin, P. Capitalization of R&D and the informativeness of stock prices. *European Accounting Review*, 16(4):703–726, 2007.
- Othman, H. B. and Zéghal, D. Analyse des déterminants de la gestion des résultats : cas des firmes canadiennes françaises et tunisiennes., 2006. URL <https://halshs.archives-ouvertes.fr/file/index/docid/581061/filename/HakimAFC2006.pdf>.
- Persson, A. and Fuentes, K. R&D Capitalization and The Income Smoothing Hypothesis – A study of Swedish listed Companies (Master Thesis). University Uppsala, Uppsala, Sweden., 2011.
- Ronen, J. and Sadan, S. Smoothing income numbers: objectives, means, and implications. Philippines: Addison-wesley publishing company inc., 1981.
- Schipper, K. Earnings management. *Accounting Horizons*, 3(4):91, 1989.
- Seybert, N. R&D Capitalization and Reputation-Driven Real Earnings Management (Partially Retracted). *The Accounting Review*, 85(2):671–693, 2010.
- Shabou, R. and Taktak, N. B. Les déterminants de la comptabilité créative: étude empirique dans le contexte des entreprises tunisiennes. *Comptabilité-Contrôle-Audit*, 8(1):5–24, 2002.
- Sougiannis, T. The accounting based valuation of corporate R&D. *Accounting Review*, pages 44–68, 1994.
- Tahinakis, P. D. R&D expenditures and earnings management: Evidence from Eurozone countries in crisis. *The Journal of Economic Asymmetries*, 11:104–119, 2014.
- Triki-Damak, S. and Halioui, K. Accounting treatment of R&D expenditures and earnings management: an empirical study on French listed compa-

- nies. *Global Business and Economics Research Journal*, 2(1):50–71, 2013.
- Trueman, B. and Titman, S. An explanation for accounting income smoothing. *Journal of Accounting Research*, pages 127–139, 1988.
- Vafeas, N. Board structure and the informativeness of earnings. *Journal of Accounting and Public policy*, 19(2):139–160, 2000.
- Vander Bauwhede, H., Willekens, M., and Gaeremynck, A. Audit firm size, public ownership, and firms' discretionary accruals management. *The International Journal of Accounting*, 38(1):1–22, 2003.
- Wang, Y., Du, R., Koong, K. S., and Fan, W. Effects of R&D policy choice on accounting performance and market value. *R&D Management*, 47(4): 545–556, 2017.
- White, G. E. Discretionary accounting decisions and income normalization. *Journal of Accounting Research*, 8(2):260–273, 1970.
- Zhao, R. Relative value relevance of R&D reporting: An international comparison. *Journal of International Financial Management & Accounting*, 13 (2):153–174, 2002.
- Zicke, J. Capitalization of R&D Costs and Implications for Earnings Management. Available at SSRN 2198307, 2014.





# Implementation of Strategic Change by Franchisees: A Sensemaking Perspective

Theresa Langenmayr

Universität Zürich

## Abstract

Franchisees play an important role when inter-organizational change, initiated by franchisors, is to be implemented. The objective of this thesis is to gain insights into franchisees' sensemaking activities. Drawing on sensemaking theories and franchising literature, a case study on franchisees in the automotive industry reveals how social processes of interaction, within and across their organizational boundaries, influence their interpretations of change initiatives. It contributes to the sensemaking literature through expanding research to inter-organizational change. It also contributes to the franchising literature through analyzing the franchisees' role during franchisor-initiated change from a sensemaking perspective.

**Keywords:** sensemaking; inter-organizational change; franchising.

## 1. Introduction

Advanced technologies, global markets and mobile capital are some of the reasons why, nowadays, continuous organizational change is essential for the competitiveness and long-term survival of companies (Lüscher and Lewis, 2008). However, managing change is one of the most challenging undertakings of an organization. The successful implementation of change can stimulate a business, whereas failure can lead to fatal consequences, including business failure (Sonenshein, 2010).

Many researchers perceive the shift in cognitions about an organization and its environment as the main purpose of strategic change (Bartunek, 1984). Following this, one of the most important processes of strategic change occurs when organizational members, when confronted with change, destroy existing and assign new meaning to an organization. This process through which individuals work to understand novel and unexpected elements that change entails is called "sensemaking" and has become a research topic of growing interest in the context of organizational theory (Maitlis and Christianson, 2014; Hope, 2010; Gioia and Chittipeddi, 1991).

This thesis draws on recent sensemaking theories. In contrast to rationalistic views on strategy, which describe strategic actions as top-down, current studies on sensemaking consider the role of actors outside the senior management team and analyze their impact on strategy formation (Balogun and Johnson, 2005). Although mainly executives design

the change, other organizational actors who operationalize new strategies serve as critical change agents. How these organizational actors assign new meaning to a system affects whether the change outcomes are in accordance or in conflict with the anticipated outcome (Hope, 2010).

In general, previous studies on sensemaking gained important insights into intra-organizational change. "Intra-organizational" implies that change is designed and implemented within an organization. Scholars focus on how change initiators and change recipients within organizations develop shared cognition, perceptions and interpretations of change initiatives (Maitlis and Christianson, 2014). Recent studies have shown that middle managers play an active role in implementing top management's change initiatives, as they mediate between the organization's strategic and operational levels (Balogun and Johnson, 2004; Balogun and Johnson, 2005; Hope, 2010).

However, little is known about sensemaking during strategic change, which goes beyond organizational boundaries. Thus, this thesis poses the research question: how does sensemaking take place during inter-organizational change?

In order to answer this research question, the thesis looks at an extreme case of inter-organizational change: the sensemaking process of franchisees during a franchisor-initiated change.

In franchising, a franchisor owns a complete business format and expands it through entering a contractual relation with independent business owners (franchisees), who pay fees for the usage of this business format (Croonen, 2010).

To align their business with upcoming threats and opportunities, franchisors need to constantly implement new strategies in their franchise systems. This can be challenging, as franchisors need to persuade franchisees to integrate these changes into their units (?). Although franchisors design the change, franchisees serve as critical change agents. As franchisees are often contractually obliged to implement new strategies and are required to make major financial investments, they tend to be skeptical towards new changes. The purposive interaction with franchisees constitutes a chance in supporting and shaping their change interpretations favorably. Thus, understanding the franchisees sensemaking activities creates useful insights for franchisors planning new strategies for their franchise systems.

A qualitative study is designed as a holistic multiple case study (Yin, 2003) conducted at car dealers in Northern Europe, where a change program, initiated by a car manufacturer, led to significant structural changes. Data was gathered from interview sessions with five car dealers, site observations and archival documents (Eisenhardt, 1989; Yin, 2003). The findings reveal how social processes of interaction influence franchisees' sensemaking activities. It is shown how and with whom franchisees mainly interact to align different possible interpretations of the change initiative.

This thesis contributes to existing literature on both sensemaking and franchising by analyzing inter-organizational change from a sensemaking perspective. Further, this thesis provides useful practical insights for franchisors, planning to implement new strategies in their franchise system.

## 2. Literature Review

The literature review consists of two parts. Most importantly, (1) the status quo of the sensemaking literature needs to be displayed in order to derive the research question for this thesis. Further, (2) a brief summary of relevant studies on franchise systems and on the franchisees' role during franchisor-initiated change serves as a basis for the case study.

### 2.1. Sensemaking

Within the next section, the literature on sensemaking is reviewed. First, the distinct characteristics of sensemaking are outlined and the sensemaking process is displayed. Then sensemaking will be applied to organizational change. The section will be concluded with a summary of studies on sensemaking of different organizational actors. The following graphic (Figure 1) serves as an overview of the sensemaking literature containing the most relevant sources. In this section, the grey marked areas will be outlined in more detail.

#### 2.1.1. Origin and definition of sensemaking

The term "sensemaking" literally refers to the making of sense. It can be understood as a process through which individuals work to grasp a novel, unexpected or confusing event

(Maitlis and Christianson, 2014). How sense is made, why, and with what effect are the primary questions motivating researchers analyzing sensemaking.

Scholars who analyzed individuals' meaning construction introduced the sensemaking language into the literature. Garfinkel (1967) first used the term "sense making" when looking into daily routines of actors as they interpret themselves and discuss with others their experience of reality. Applying sensemaking to the field of organizational theories, Salancik et al. (1977) studied how an individual's current action limits his future choices and actions. In the 1980s, researchers concentrated more on the cognitive aspect of sensemaking, developing the main steps of the sensemaking process (Louis, 1980; Milliken, 1987). In the 1990s, the research on sensemaking broadened and sensemaking related concepts were introduced (Gioia and Chittipeddi, 1991).

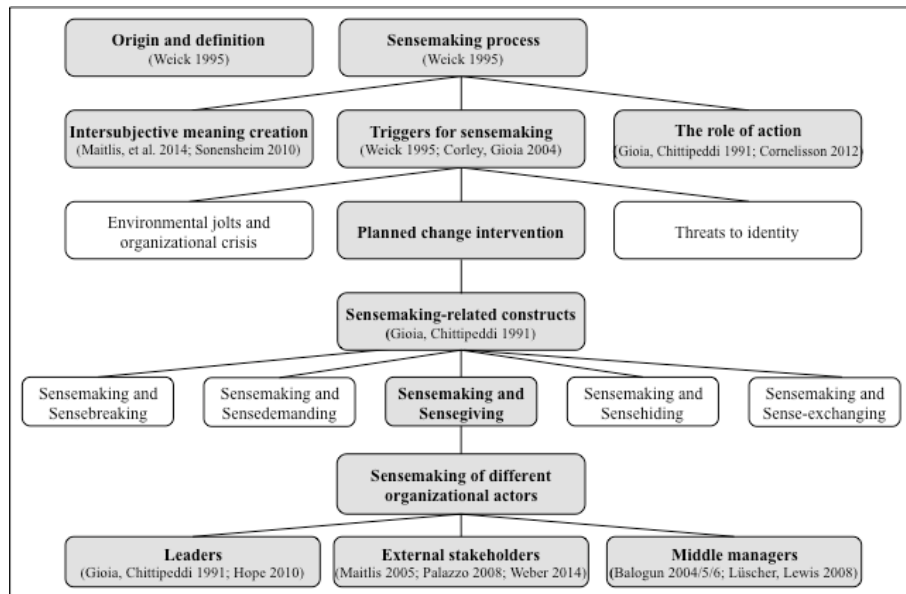
Due to the high number of studies conducted, there is no universal "sensemaking theory", but rather diverse concepts of sensemaking (Weick, 1995; Balogun and Johnson, 2005; Maitlis, 2005). Hence, Weick identifies seven distinguishing characteristics of sensemaking, which are often mentioned in literature and distinguish sensemaking from other explanatory processes (Weick, 1995, p. 17-61).

First of all, sensemaking is self-referential as it depends on the person who engages in the sensemaking. The person is rather making sense about himself than about the environment, determining the implications for his personal situation. The process of interaction constitutes an individual. Thus, the sensemaker is continuously reevaluating how he should present himself to others and how he perceives himself. The more perceptions a person has of himself, the more meaning needs to be extracted and imposed in any given situation (Weick, 1995, p. 18-24).

The second characteristic describes sensemaking as retrospective, which means that sensemaking is affected by the individuals' "lived experiences" (Schutz, 1967). Louis (1980) describes sensemaking as a retrospective thinking process. During that process, individuals form unconscious and conscious expectations about future events. When the actual event differs from previous predictions, meaning needs to be assigned as an output of the sensemaking process. As various meanings could be assigned in retrospective sensemaking, the individual needs values, priorities and clarifications on preferences as a basis for prioritization. Thereby, emotions affect sensemaking because they influence which retrospective activities are considered.

Weick (1995) calls the third characteristic "enactive of sensible environments", which refers to the actual "making" of sense. By assigning meaning to a situation, people constitute the environment that they seek to understand. Thus, enactment is based on the idea that people play a key role in creating the environment in which they find themselves. "People act, and in doing so create the materials that become the constraints and opportunities they face" (Weick, 1995, p. 30).

Fourthly, sensemaking is social as individuals, who involve themselves in sensemaking, are always part of a socio-



**Figure 1:** Overview sensemaking literature; Source: Own figure

material context. People negotiate and create a shared sense of meaning with others (Balogun and Johnson, 2004). Maitlis (2005) describes sensemaking as a social process in which “organizational members interpret their environment in and through interactions with each other, constructing accounts that allow them to comprehend the world and act accordingly”. Although sensemaking in organizations may either take place through horizontal communication (within a hierarchical level) or as a result of top-down communication, it is mainly determined by the social processes at the recipient level (Balogun, 2003). Further, sensemaking is social as organizational decisions are made either mutually or with the knowledge that individual decisions impact other organizational members (Weick, 1995, p. 49).

The fifth characteristic describes sensemaking as ongoing. Sensemaking has no starting point, such that meaning is made “in an ongoing present in which past experience is projected upon possible futures” (Hernes and Maitlis, 2010). The reality is understood as a flow of continuous activities, which is interrupted by events that trigger a need for explanation. Through sensemaking meaning is extracted and assigned to this events, and thereby focal points for future streams of activity are accomplished. Emotion is what happens between a stream becoming interrupted and the interruption being removed.

Referring to the sixth characteristic, Weick (1995) says that sensemaking “focuses on and by extracted cues”. When individuals are confronted with a discrepant event that is surprising, single elements are extracted to construct the character of the whole event. In order to decrease the complexity of reality, they seek to clarify what is going on by extracting and interpreting these elements, so-called cues, from the environment. Research concentrates on ways individuals notice and extract cues. Hereby, the context affects what is extracted as

a cue and how it is interpreted. Furthermore, as sensemaking is retrospective, past experience dictates what cues are chosen (Helms Mills et al., 2010). These cues serve as a basis for a simplified, subjectively perceived logical order that “makes sense” of what has occurred. This logical order is fundamental and contains consistent sets of organizing principles, which are unquestioned (Ford and Ford, 1994). These formed logics enable actors to cope with a complex reality. The control over which cues serve as a point of reference is an important source of power.

Lastly, the seventh characteristic describes sensemaking as driven by plausibility rather than accuracy. This characteristic implies that individuals do not rely on the accuracy of their perception but rather look for cues that make their sensemaking seem plausible (Helms Mills et al., 2010).

After having elaborated the seven distinct characteristics of sensemaking, the different phases of the sensemaking process will be described.

### 2.1.2. Sensemaking process

The sensemaking process can be divided into three main steps: First, the cues of an unexpected interruption are perceived (triggers for sensemaking). Then interpretations for this interruption are created (inter-subjective meaning creation), followed by actions taken (the role of action on sensemaking).

As previously mentioned, sensemaking occurs when salient, novel, unexpected or confusing cues are noticed that violate perceptual frameworks (Maitlis and Christian, 2014). However, not all unexpected events trigger sensemaking. Sensemaking occurs when the discrepancy between previously formed expectations and actual events is great enough and important enough to provoke individuals to ask what is going on and how they should further proceed. Such discrepancy occurs when either the new event

is not expected (does not fit into the ongoing interpretation of the environment) or an expected event does not happen. While distinguishable, these two types lead to similar consequences. The interruption consumes information-processing capacity and therefore forces individuals to concentrate on cues that can be used in sensemaking. The loss of cues leads people to focus their attention on those aspects of the interruption judged most important.

Sensemaking is further triggered by cues for which the meaning is not only judged important but also ambiguous or for which the outcome is uncertain. Ambiguity and uncertainty are occasions that are prominent in organizations. The “unexpected” in each of the two cases is somewhat different. “Ambiguity” means the presence of two or more possible interpretations of an event. These interpretations consist of different elements such as perspectives, feelings, messages, demands, identities, interest or practices, which become evident through social reflection or interaction (Lewis, 2000). Putnam (1986) describes in his study three different sources of ambiguous tensions. Within self-referential loops (1) ambiguity is embedded in a cohesive statement or concept. An example is the circular statement “I am lying”. (2) Mixed messages describe inconsistencies between verbal and non-verbal responses during social interactions. One example would be an organization that supports teamwork while monitoring the performance of employees on an individual basis. The third source of ambiguity are system contradictions (3), referring to objectified tensions within systems, such as organizations consisting of autonomous teams and formalized processes at the same time (Clegg et al., 2002). Reasons for growing ambiguous conditions are increasing information, complexity, turbulence and competition (Cameron and Quinn, 1988, p. 3). Therefore, ambiguity is perceived when a lack of clarity and high complexity make multiple explanations of a novel event possible. The problem in ambiguity is not that information on the real world is lacking, but that information may not resolve misunderstandings (Weick, 1995, p.93). Thus, sensemaking is triggered through the existence of too many possible interpretations, whereas in the case of uncertainty, individuals engage in sensemaking because they are not capable of forming any interpretations (Burns and Stalker, 1961). If an individual is uncertain, it is not a problem of too many interpretations, but of too few. Milliken (1987) states that the definition of uncertainty has three aspects: (1) people lack understanding of how components of the environment are changing (state uncertainty), (2) of the impact of the environmental changes on the organization (effect uncertainty), or (3) of the response options that are open to them (response uncertainty).

Besides ambiguity and uncertainty as general triggers for sensemaking, studies have explored different contextual triggers for sensemaking, such as environmental jolts, organizational crisis, threats to organizational identity and planned organizational change initiatives. Organizational change, in contrast to the other described triggers, is anticipated and planned. It violates expectations of organizational actors and generates both uncertainty and ambiguity. Change pro-

cesses can either directly target organizational meanings or may begin with a structural transformation that violates existing perceptions of the organization and therefore leads to sensemaking. Despite the structural transformation itself, the sensegiving of leaders initiating the change can also trigger sensemaking (Maitlis and Christianson, 2014).

Now that possible triggers for sensemaking have been identified, the inter-subjective meaning construction will be looked at in more detail. In most of the studies, sensemaking is understood to be concerned with language. These studies can be divided into those which highlight the importance of narratives, of metaphors or the local and situated nature of discursive practices (Maitlis and Christianson, 2014). The most important findings of these studies will be further outlined below.

A narrative can be described as “the primary form by which human experience is made meaningful” (Polkinghorne, 1988). It is a verbal or written story about events whose goal it is to create meaning through putting new experience in already established categories. Maitlis and Christianson (2014) perceive narratives as a benefit to organizational research because a narrative reveals not only an individual’s involvement and actions but also the process through which meaning is formed. Sonenshein (2010), who conducts a study on strategic change, claims that organizational members construct shared narratives. In analyzing the process through which this happens, Sonenshein argues that although leaders may want to convince employees of their reality through sensegiving, employees can use the leaders narratives to construct their own meaning. He shows that employees adopt management’s narratives about the organization, but draw them on their own experience. This could result in either a supportive or a subversive narrative.

Metaphors, as a figure of speech identifying similarities between concepts, are used to connect cues and frames (references based on past experience). Weick (1995) says that metaphors have the ability to create order in unfamiliar situations, to evaluate and to provide justification for certain actions. Gioia and Chittipeddi (1991) show how organizational members use metaphors during shared sensemaking activities first to build a mutual social identity and later to legitimize their role in the change initiative by redefining themselves. Cornelissen (2012) further analyzes the use of metaphors in sensemaking and claims that the use of metaphors is influenced by the individuals’ commitment to fulfill certain tasks as part of their professional role and others’ expectations about them. Thus, the sensemakers’ use of metaphors depends on the interruptive event and the socio-cultural context (Maitlis and Christianson, 2014).

In diverse studies Balogun et al. focus on discursive practices to highlight the importance of the socio-cultural context for meaning construction in organizations. In the social process of sensemaking, change recipients interact in exchanging stories, rumors and conversations (Balogun and Johnson, 2005). When analyzing how middle managers engage in sensemaking, Rouleau and Balogun (2011) explore their formal and informal conversations with relevant actors in-



side and outside the organization. They describe, “setting the scene “ as decisive - a set of practices used to create the context of the conversation. The most important practices are the setup of an organizational network and the strategic involvement of the right people. Rouleau and Balogun (2011) conclude that middle managers need an understanding of the context as a basis for the appropriate use of language and attitudes in order to connect to different actors. Sensemaking and the influence of others are thus not only accomplished through language, but through the sensitization to its situated context.

The last part of the sensemaking process is the sensemakers’ action taken. The sensemakers’ actions are an integral part of sensemaking as the observation of outcomes and the consequential learning can serve as input for additional sensemaking activities. Generally speaking, the sensemakers’ inter-subjective meaning construction can either lead to a defensive reaction or a proactive management of the novel event. Vince and Broussine (1996) catalogue five different types of defensive reactions. A repression (1) is blocking unpleasant experiences from memory. Through the application of a regression (2), the individual resorts to actions that have provided some security in the past. A projection (3) helps to transfer personal shortcomings to others, so bad feeling are transferred into a repository. Reaction formation (4) entails manifesting the feeling or practice opposite to the threatening one. In the case of denial (5), the individual refuses to accept an unpleasant reality.

Tapping the positive potential of a novel event means finding a way of managing it. Lewis (2000) conducts a study on sensemaking triggered by organizational paradoxes. Paradoxes are a special form of ambiguity, where “contradictory, mutually exclusive elements are present and operate equally at the same time” (Cameron and Quinn, 1988, p. 2-3). According to Lewis (2000), managing paradoxes during change means capturing its enlightening potential. The target is to rethink and reframe past perceptions and practices. Three means of managing paradoxes are mentioned. (1) The individual can accept the paradox and learn to live with it. (2) Another possibility is to confront oneself with the paradox; discussing tensions to socially construct a more accommodating understanding. (3) The individual can use its mental capacity to think paradoxically. Through the application of reframing (reconsidering previously established frames), tensions become viewed as interwoven.

The possible actions taken by the sensemaker are closely linked to the previously described sensemaking characteristic “enactive of sensible environments”, which refers to the reciprocal relation between the sensemakers’ action and his environment. According to Maitlis and Christianson (2014), action is important for sensemaking for three reasons. First, they create the input for sensemaking by generating stimuli and cues from which one is able to learn about situations. After taking actions one can observe the cues generated by the action. Second, actions can test the understanding generated through prior sensemaking activities. Thereby, actions serve as input for new sensemaking, while simultaneously provid-

ing feedback about the sense that has already been made. Thirdly, actions shape the environment for sensemaking and therefore change the situation that initially triggered sensemaking.

Figure 2 visualizes the sensemaking process.

### 2.1.3. Sensemaking and sensegiving

As the amount of research on sensemaking increases, so does the number of specialized forms of sensemaking and sensemaking-related constructs. Examples of specific forms of sensemaking are “cultural sensemaking”, “interpersonal sensemaking” and “ecological sensemaking”. The most popular sensemaking related constructs are “sensebreaking”, “sensedemanding”, “sense-exchanging”, “sensegiving”, “sensehiding” and “sense specification” (Maitlis and Christianson, 2014). For this thesis, sensegiving will be further outlined and applied to the case study in a later section.

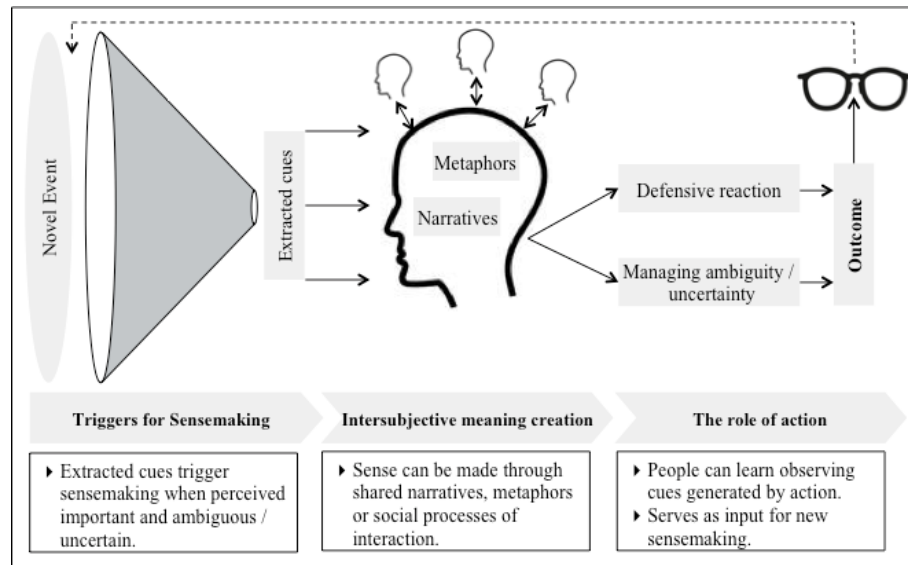
Sensegiving is often studied in the context of how organizational leaders try to strategically influence the sensemaking of other organizational members. Polanyi (1967), who first introduced the concept, uses sensegiving to describe how people fill speech with meaning and make sense of speech. Gioia and Chittipeddi (1991) conduct a decisive study on sensemaking and sensegiving in organizational change. They describe sensegiving as “the process of attempting to influence the sensemaking and meaning construction of others toward a preferred redefinition of organizational reality”. They develop a sensemaking-sensegiving-process during which not only leaders, but at a later point also other organizational actors, become involved in sensemaking and sensegiving as a response. Thus, sensegiving is not solely regarded as top-down, as the recipients of leaders’ sensegiving activities can form their own interpretations and can even resist or reject the sense they have been given. At a later point of this thesis it will be described how an organization, which has designed a change initiative, undertakes sensegiving to influence the sensemaking of an actor responsible for operationalizing the strategy.

### 2.1.4. Sensemaking and organizational change

In many studies sensemaking is seen as an explanatory mechanism for different organizational processes. The impact of sensemaking on organizational processes, such as strategic decision-making, innovation and creativity as well as organizational learning, has been analyzed. For this thesis, studies on the relation between sensemaking and organizational change will be further outlined.

According to Ford and Ford (1994), change involves two interrelated elements. First, the identity of what something is. And second, the change process as the movement from old to new. Internal ambiguity arises when new practices are incompatible with old practices during the phase of organizational change. Then, conflicts between incompatible ways for conducting business evolve. According to Westenholtz (1993), it is “the interplay of internal contradictions that brings about the change”. Westenholtz states that organizational sensemaking occurs when employees experience an





**Figure 2:** Sensemaking process; Source: Own figure

ambiguous situation in the organization. When paradoxes, as an extreme form of ambiguity, occur during organizational change, researchers speak of “dialectic change”. In dialectics a “thing” is a unity of contradictory opposites. “The aim of dialectical inquiry is to improve decision quality by purposefully infusing tensions into group dynamics, then seeking resolution by synthesizing divergent alternatives” (Lüscher and Lewis, 2008).

The relation between organizational change and sensemaking is recursive, as change can trigger sensemaking and sensemaking can also accomplish strategic change. Managers (independent of the organizational level) can apply sensegiving to convince others of the value of changes so they perceive the redefined reality the same way (Maitlis and Christianson, 2014). When managers successfully influence the sensemaking of other organizational actors, these individuals are motivated to make changes in their own roles and are able to help others to co-construct ways of working that are consistent with it (Gioia and Chittipeddi, 1991). Thus, actors create a new organizational order through sensemaking in response to environmental changes.

#### 2.1.5. Sensemaking of different organizational actors during change

For a more comprehensive understanding of sensemaking during organizational change, researchers analyze the sensemaking process of different organizational actors. Literature highlights the importance of sensemaking of both leaders and other organizational actors during organizational change. Studies that analyze the role of leaders found that organizational sensemaking is often limited with leaders controlling the process. When leaders decide to integrate other organizational actors in the change process, they rather target the creation of legitimacy than an adaptation of the initiative based on employees’ feedbacks (Gioia and Chittipeddi, 1991). Based on Gioias’ and Chittipeddis’

sensemaking-sensegiving process (1991), other studies further examine the varieties of leaders’ sensegiving. Gioia and Thomas (1996) find that sensegiving depends on the political nature of an issue. Bartunek et al. (1999) derive different leaders’ sensegiving strategies. Within one of these strategies it is shown that leaders who feel personally threatened are more likely to focus their sensegiving efforts upon the opportunities that a strategic change entails.

In contrast to sensemaking controlled by leaders, sensemaking can also appear uncontrolled when neither the members’ sensemaking activities are organized, nor the outcomes are integrated into a collective account. Balogun and Johnson (2005) perceive middle managers as decisive change agents, who are often responsible for implementing new structures, but have no impact on the upfront change design. The authors study sensemaking of middle managers during a top-down change initiative in which the structure of an organization becomes more decentralized. The unexpected restructuring creates ambiguity within the middle managers and thus triggers sensemaking. The findings suggest that senior managers cannot directly influence the sensemaking of middle managers, but affect organizational meaning creation through the presence of their actions in rumors and stories shared by others. The authors describe sensemaking as a process of social interaction (Balogun and Johnson, 2004).

In their following study, Balogun and Johnson (2005) focus on these processes of social interaction to better grasp how middle managers try to make sense of change interventions. As previously mentioned in the description of the sensemaking process, they show the extent to which lateral, informal social processes and inter-recipient sensemaking influence the change outcome. As middle managers try to implement the change plans of their seniors, their everyday experiences of this implementation and the related behaviors of others, and the stories, gossip, jokes, conversations and

discussions they share with their employees about these experiences, shape their interpretations of what they should be doing. "Change interventions and plans are translated into action through the medium of these inter-recipient processes, turning top-down intended change into an emergent and unpredictable process".

Based on the findings of Balogun and Johnson (2004), Balogun and Johnson (2005) and Hope (2010) conducts a study on the impact of organizational politics on middle managers' sensemaking and sensegiving activities. The objective is to identify the politics used by middle managers to influence the change outcome. His findings suggest that middle managers have an influence on the sensemaking of seniors by exerting the "power of meaning". Thereby, the author refers to middle managers' resource power, such as special expertise in a business unit, used as a tool for controlling the change process.

Lewis (2000) apply organizational paradoxes to sensemaking when analyzing the role of middle managers in the process of change. They understand sensemaking as a form of reframing, which enables actors to alter meanings attributed to changing situations. Their findings show how organizational change aspects, such as paradoxes of performing, belonging, and organizing, can be transformed "from a label to a lens" through the process of working through paradox. This process of working through paradox contains five parts: At first, middle managers are confronted with an intricate, fluid, and fuzzy issue the authors call "mess" (1). Actors initiate sensemaking by narrowing down this mess to what is of interest to them. Thereby, they set boundaries for exploration. The outcome is the definition of a more specific problem (2) to enable reflection. A clearly stated problem places an issue on an agenda for solution finding. In the next step, they analyze why managers felt incapable of solving problems through the identification of more complicated, underlying dilemmas (3). A dilemma, in contrast to a paradox, is a form of ambiguity that requires managers to make an either-or decision between polarities. Uncovering opposite sides of an issue lets managers feel more paralyzed. In this context, paradoxical thinking (4) let the managers recognize how the tensions between opposing sides are needed and interwoven. "As managers moved from a mess to a problem to a dilemma to a paradox, each stage encouraged deeper exploration toward a more workable certainty" (5). Workable certainty implies that managers are always in the process of sensemaking. Lewis (2000) conclude that working through paradox aids sensemaking because paradoxes serve as a mean for managers to consider different perspectives, alter their assumptions and explore issues in fundamentally different ways. Thus, through paradoxical thinking the sensemaker is able to accommodate, rather than eliminate, persistent tensions.

Although most of the research on sensemaking focuses on actors within the organization, some studies consider external actors during organizational change. Basu and Palazzo (2008) develop a process model of sensemaking through studying a company, which decided to engage in corporate

social responsibility (CSR). The authors explain how managers think, discuss and act with respect to internal and external stakeholders. The CSR strategy was formulated not as a direct response to external demands but, instead, through cognitive and linguistic sensemaking.

Weber et al. (2015) apply the sensemaking-sensegiving process of Gioia and Chittipeddi (1991) to external stakeholders. The authors want to understand the interactions between internal and external stakeholders during a post-9/11 organizational change in the Maritime Transportation Security Act. The external stakeholders considered are the government and the society who need to legitimize the change. They identify four critical triggers that lead to a communication breakdown in the divergent sensemaking process: (1) unidirectional and lacking communication, (2) multifaceted understandings of organizational identities, (3) misaligned cues and (4) the emergence of inter-organizational sensemaking. Their findings illustrate the importance of considering a broad perspective of legitimate participants in a sensemaking process, as well as constantly adjusting sensemaking activities to avoid contradictions between participants' perspectives.

During this section, relevant research findings on sensemaking have been outlined. Various studies already gained profound insights into the sensemaking process of different organizational actors. However, studies on sensemaking during change only regard intra-organizational change. But what if an organization decides to implement a self-planned change initiative outside its organizational boundaries? The change implementer then becomes an important change agent, yet his sensemaking process is currently under-researched. Therefore, this thesis poses the following research question: How does sensemaking take place during inter-organizational change?

The focus of this thesis lies on the sensemaking process of the actor, who implements the change, and not on the organization, which designs the change.

In order to answer the research question, a case study is conducted using an extreme case of inter-organizational change: the sensemaking process of a franchisee during a franchisor-initiated change. This case is extreme, as the franchisee is contractually obliged to implement changes. Further, he needs to invest capital himself and trust in the profitability of the change without being able to influence the upfront design.

To analyze the sensemaking of franchisees in the process of franchisor-initiated change, mainly findings of studies on the sensemaking of middle managers are applied. Neither middle managers nor franchisees design the change initiative. Nevertheless, both have the task of operationalizing the strategy in their business unit and therefore play a decisive role in the change process. However, the roles of middle managers and franchisees have a central distinction: the middle manager is an internal while the franchisee is an external stakeholder. Hence, different interpretations and possible contradictions arise when they are confronted with unexpected change.

In order to better grasp the franchisor-franchisee relationship and the organizational role franchisees entail during inter-organizational change, the following section outlines related literature.

## 2.2. Organizational change in franchising

Franchising is an important business format, constituting up to 50% of retailing sales in Western countries (?).

“In franchising, an organizing firm (franchisor) enters into a contractual relationship with franchisees –typically small business owners who pay to use a business format and agree to conform to franchisor standards” (?). On the one hand, a franchise system provides a solution for certain agency problems, since franchisees are meant to be more motivated than company managers running independent business units that have financial investments at stake and receive unit profits. On the other hand, problems may arise in the franchisor-franchisee relation (Dada et al., 2012).

A growing number of studies in management literature deal with franchising, mainly focusing on difficulties in managing franchise relationships (Davies et al., 2011). Especially the tension between the franchisors’ wish for system standardization and the franchisees’ wish for entrepreneurial autonomy is of interest. Franchise system standardization can be understood as the level of obligations the franchisee has to fulfill when running his unit, and the degree to which these obligations are monitored by the franchisor (Cox and Mason, 2007). Thus, a higher degree of standardization leads to a higher franchisors’ decision-making authority.

The aforementioned tension between the franchisors’ wish for system standardization and the franchisees’ wish for entrepreneurial autonomy becomes even higher when the franchisor decides to implement transformational change forcing the franchisee to make a significant investment in their unit. Introducing change into a franchise system can be more challenging than within hierarchical organizations, as independent business owners need to be convinced of the opportunities a change initiative entails for their own unit. While many changes imply financial investments, the franchisee needs to trust in the profitability without being able to influence the change design. Under these circumstances, the relation between franchisor and franchisee may be put under increased pressure (?).

However, only a few studies deal with the role of franchisees during strategic change although their response is key to a successful implementation. Croonen (2010) generates a theory on franchisees’ perceptions of trust and fairness during change. She provides explanation for franchisees’ perception of distrust and unfairness resulting in resistance or destructive behavior. She shows that the franchisees’ trust is mainly based on his perception of distributive, procedural, and interactional fairness. In order to create, maintain or increase franchisees’ trust, the franchisor should have company-owned units and should integrate franchisees’ feedback into the strategic decision-making process.

Later, ? conduct a qualitative study to grasp the franchisees’ reaction to strategic change. The authors identify

three groups affecting franchisee responses and response-switches during franchisor-initiated change; “the different dimensions of expected franchisees satisfaction, a franchisees’ trust in the franchisor and the perception of fairness of the change process as well as the franchisees’ perception of alternative attractiveness and switching costs”. Their findings reveal that expected satisfaction regarding the profitability and trust in the franchise system were the most important reasons for franchisees to adopt a constructive or destructive response or to switch between them. Further the authors can show that franchisees who perceive a higher level of standardization, take more dimensions of trust and satisfaction into account when evaluating the franchisor and the franchise system. When franchisees have a high level of trust and satisfaction, they tend to lower the importance of economic motivations such as alternatives or switching costs.

Although these studies help to gain insights into franchisees’ reactions to franchisor-initiated change, the process through which franchisees work to understand change remains unexplored. This thesis applies the concept of sense-making to the field of organizational change in franchising to fill this research gap. For the upcoming case study, the following research question is posed: How does sensemaking take place at franchisees during franchisor-initiated change?

## 3. Research Design

Eisenhardt (1989) developed a process model to build theory from case study research, which serves as a framework for the research design of this thesis. This process ensures that the so-called “chain of evidence” is maintained, which implies that an external observer should be able to follow the derivation of any evidence, ranging from initial research question to case study conclusion (Yin, 2003, p. 105). Eisenhardt’s research process aligns and structures different research approaches. The most prominent approaches are grounded theory (Glaser and Strauss, 1967), which relies on continuous comparison of data and theory, Yin (2003) design of case study research and Miles and Huberman (1984) techniques for analyzing qualitative data. Further literature is considered across the process stages.

Figure 3 summarizes the process stages, which will be further outlined in the following sections.

### 3.1. Getting started

#### 3.1.1. A priori research and deduction of research question

According to Eisenhardt (1989) the first step when building theory from case study research is the definition of a research question and the building of an a priori construct. Within the first two chapters of this thesis, the theoretical framework was constructed and the research question derived. The views on how much research should be conducted in advance differ. Glaser and Strauss (1967), who introduced grounded theory into literature, share the opinion that a priori constructs limit the researcher’s openness to exploratory

|                                                                                                                                                                                                                            |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>1. Getting started</b>                                                                                                                                                                                                  |
| - Definition of research question<br>- A priori specification of construct through review of theory (Miles, Huberman 1984)                                                                                                 |
| <b>2. Selecting cases</b>                                                                                                                                                                                                  |
| - Holistic multiple case design (Yin 2003, p. 40)<br>- Theoretical sampling (Glaser, Strauss 1967, p. 45), environmental variation (Pettigrew 1990)                                                                        |
| <b>3. Crafting instruments and protocols</b>                                                                                                                                                                               |
| - Data collection according to triangulated method (Pettigrew 1990): interviews, observations, documents<br>- Research protocol, field notes (Yin 2003, p. 87)                                                             |
| <b>4. Entering the field</b>                                                                                                                                                                                               |
| - Overlap data collection and analysis (Glaser, Strauss 1967)<br>- Flexible and opportunistic methods → adding questions or adopting data collection method (Pettigrew 1990)                                               |
| <b>5. Analyzing the data</b>                                                                                                                                                                                               |
| - Within case analysis (descriptive write-ups) and cross case analysis (constant comparative model)<br>- Iterative tabulation of evidence for constructs (Glaser, Strauss 1967), Logic across cases (Miles, Huberman 1984) |
| <b>6. Enfolded literature</b>                                                                                                                                                                                              |
| - Comparison with conflicting literature (Eisenhardt 1989)<br>- Comparison with similar literature (Eisenhardt 1989)                                                                                                       |
| <b>7. Reaching closure</b>                                                                                                                                                                                                 |
| - Theoretical saturation when possible<br>- Final report                                                                                                                                                                   |

**Figure 3:** Overview research process; Source: Own figure in accordance with Eisenhardt, 1989

findings. However, as there have already been many studies published in both the fields of sensemaking and franchise systems, the findings have been considered in the formulation of a relevant research question for this thesis. Furthermore, Eisenhardt (1989) and Yin (2003) claim that a priori constructs provide a better grounding of construct measures. However, although a theoretical framework has been established, it has been avoided to develop concrete hypotheses from pre-existing theories prior to the data collection. Pre-conceived hypotheses structure both data collection and analysis in advance and limit developing ideas, which are essential for exploratory research (Charmaz, 1990).

### 3.1.2. Determination of research strategy

Social science research can be conducted in several ways including experiments, surveys, histories, analysis of archival information and case studies (Yin, 2003, p. 1). When deciding for a research strategy and its own logic, Yin (2003) names three conditions to consider: (1) The type of research question posed, (2) the extent of control an investigator has over actual behavioral events and (3) the degree of focus on contemporary as opposed to historical events. This thesis poses an exploratory research question, which is a justifiable reason for using an exploratory study, which has the goal of developing, instead of testing, hypotheses and propositions for further inquiry (Yin, 2003 p. 6). Related to the second and third mentioned condition, this thesis wants to examine a temporary event, where the relevant behavior cannot be manipulated. Thus, the case study is an appropriate exploratory research method to choose. Other explanatory methods build on different data sources. History, for instance, is a method to deal with a past occasion where no primary resources are available. Experiments are done when the behavior of participants can be manipulated.

As the case study is chosen as an appropriate research strategy for this thesis, the distinct characteristics will be briefly outlined. First, a case study examines a current occasion within its real-life context, particularly when the boundaries between occasion and context are not clearly obvious. Second, the case study benefits from prior theoretical propositions and deals with occasions in which many more variables are of interest than data points could be considered (Yin, 2003, p. 14-15).

There are four different types of case study designs, depending on how many different contexts and units of analysis are integrated into the case study. First, the researcher has to decide whether he is using a single or multiple case-study design, depending on the number of cases considered. Possible reasons for choosing a single case are that it represents a critical case in a well-formulated theory or that it is an extreme or unique case. Second, the researcher has to determine if he either integrates a single-unit of analysis (holistic) or multiple units of analysis (embedded) into his study.

For this thesis a holistic multiple-case study design is used. In general, multiple cases enable replications so that the evidence is regarded as being more robust compared to single-case studies. However, although no single rare or critical case is regarded as in the single-case study design, the cases chosen still need to fit for a particular purpose within the complete spectrum of research. The cases can be selected either because “they predict similar results (literal replication) or because they predict contrasting results for predictable reasons (theoretical replication)” (Yin, 2003, p. 47). Based on the theoretical framework and the decision for a single-unit analysis, this thesis aims for a theoretical replication. These replication logics specifically apply to case studies and cannot be compared to quantitative sampling methods where statistical methods are used to obtain results



for subgroups (Yin, 2003, p. 48).

### 3.2. Selection of cases

As previously mentioned, the multiple-case study design requires a justified selection of cases. Similar to the field of theory-testing, the selection of cases in theory-building is decisive, as the cases chosen determine the set of entities from which the research sample is to be built (Eisenhardt, 1989). Glaser and Strauss (1967) introduce the concept of theoretical, in contrast to random, sampling. The basic question in theoretical sampling is: "What groups does one turn to next in data collection and for what theoretical purpose?" (Glaser and Strauss, 1967, p. 47) As apparent in this question, the criteria of choosing groups in theoretical sampling are those of theoretical purpose and relevance – not of structural circumstances (Glaser and Strauss, 1967, p. 48). The researcher chooses any cases that will help to extend theory. As mentioned before, the purpose of exploratory studies is to generate theory, not to validate "facts". Further, theoretical sampling provides for ongoing joint collection and analysis of data. The researcher can decide in the process of data collection to align the cases considered as unexpected findings may occur. The ideal situation according to grounded theory is that the number and types of groups can only be cited when the research is completed.

Group comparison within a case study is used to compare divergent or similar evidence indicating the same conceptual categories and properties (Glaser and Strauss, 1967, p. 49). As previously mentioned, this thesis applies the holistic multiple-case study design. This is also due to the fact that more than one case allows for comparison groups.

The control over similarities and differences helps in discovering categories when analyzing the data. On the one hand, minimizing the differences among cases increases the chance of collecting a high amount of similar data on a given category, and of spotting important differences not caught in previous studies. On the other hand, maximizing the difference among cases increases the probability of collecting heterogeneous data on a category (Glaser and Strauss, 1967, p. 55-56).

As time available for obtaining results for the thesis is limited, the groups are chosen at the beginning of the data collection. However, the theoretical framework helps to extend the theory on sensemaking and franchise systems to the not-yet-analyzed inter-organizational change. Thus, in accordance with grounded theory, the sample is not random but reflects the selection of specific cases to extend the theory. The idea of minimizing differences is applied to the case selection.

The chosen cases will be further outlined in the following section.

#### 3.2.1. Case study setting

The case study sites are dealers of a car manufacturer. The car manufacturer has almost 5,000 dealers worldwide and an annual sales volume of over 2 million cars in 2014 (Anon.).

In order to get a better understanding of the organizational role car dealers entail, Figures 4 and 5 illustrate how the dealer is embedded in the sales process and how the contractual relation between the car manufacturer and the dealer can be characterized.

#### 3.2.2. Introduction to change initiative

The car manufacturer initiated a strategic sales initiative in 2010. Target of this change program is to enhance the point of contact to the customer. Therefore, most of the centrally-designed concepts are rolled out at the dealerships. The development and implementation of the change initiative leads to a globally-standardized sales process at the dealerships.

In order to obtain a standardized premium customer experience at car dealerships, different concepts are holistically implemented under the umbrella of this change initiative. The most prominent concept is the Product Genius, who is acting as a product specialist, matching the customers' demands with the product features. The Product Genius is introduced into the dealerships as a new role among the existing sales staff. The customer relations of sales persons change, as the Product Genius is now responsible for explaining technical features to the customer without having a financial incentive to sell a car.

In addition, the Customer Treatment is standardized through training and coaching of the staff at all levels of the dealerships. Thus, the individual and self-determined relation of the sales persons to the customer is restricted.

The use of iPads, including an app (Mobile Customizer) for configuring the car with the customer, and a screen (Virtual Product Presenter/VPP) for mirroring the configured car, shall further improve the sales process.

Besides the standardization of the sales process including the use of new IT tools, the change initiative targets a consistent visual appearance of car dealerships; thus, the car manufacturer sells their dealers centrally-purchased premium furniture (EPoS). The different and more open arrangement of the workplace changes the customer relation as well.

In order to monitor the standardization of the sales process and the fulfillment of the new Retail Standards, the car manufacturer conducts Mystery Shopping at the dealerships. The execution and, for most of the dealers, the results of the Mystery Shops affect their dealer bonuses.

Implementation managers in the Headquarter and National Sales Companies, as well as an external agency, who is conducting the training and coaching modules, monitor the implementation of the change initiative at the dealerships. The implementation process within one market or region takes approximately two years, depending on the number of dealerships. The phases of the implementation are pre-determined. Managers at the dealership are informed and coached first, so they can serve as on-site change agents. Extensive change and communication campaigns support the implementation managers at the National Sales Company in explaining the targets and benefits of the change initiative to the dealers.



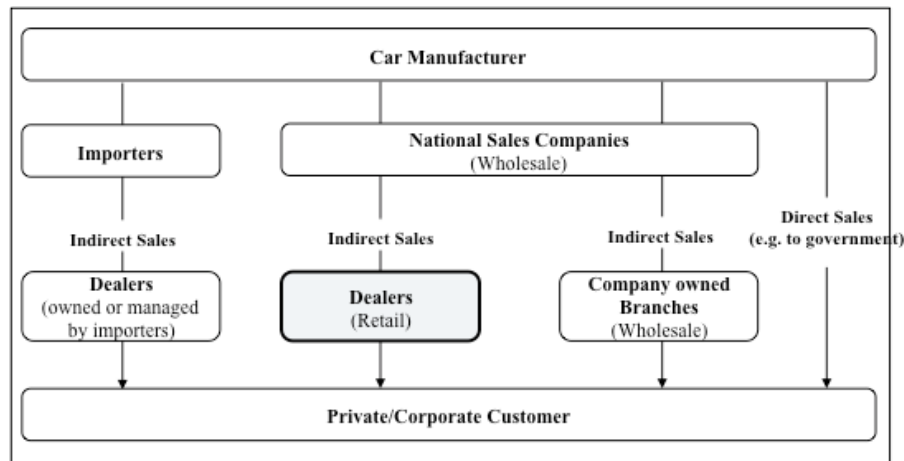


Figure 4: Sales process of car manufacturer; Source: Own figure

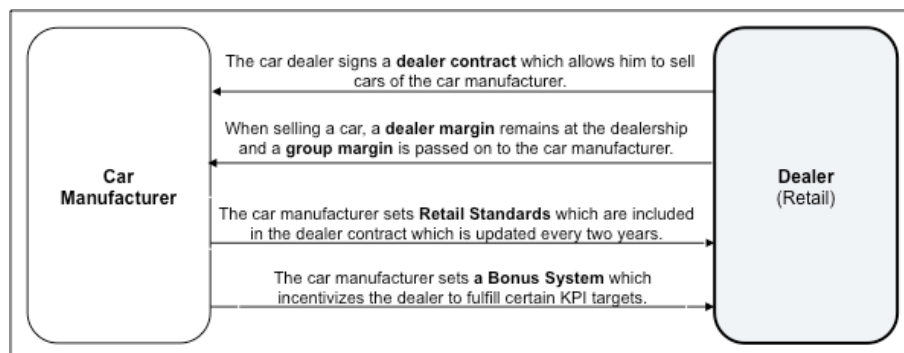


Figure 5: Contractual relation of dealer and car manufacturer; Source: Own figure

Besides financial support from the car manufacturer for the initial funding, the dealers pay most of the implementation cost. Further, most concepts of the change initiative are part of the Retail Standards and the bonus system (see Figure 4), which makes the implementation of these elements mandatory for the dealer.

Overall, the project holds many benefits for the dealer. However, their autonomy in designing the sales process at their dealership is restricted through the implementation of the change initiative. Confronted with this new situation, the question arises at what point do ambiguity and uncertainty, which serve as triggers for sensemaking, are generated at the dealer.

### 3.2.3. Case study sampling

Car dealers, who can own one or more dealerships, are selected as the recipient group for the research of this thesis since they play a key role in the implementation of the aforementioned change initiative, yet their role (franchisees that implement franchisor-initiated change), remains under-researched in the literature. The new strategy requires the dealers to undertake significant structural change at their dealership(s). Similar to middle managers, dealers can be the target and agents of change (Balogun and Johnson, 2005). Further, they have to implement the plans devised by

the Headquarter and National Sales Companies, with little involvement in the upfront decision-making.

The primary method for collecting data is focused interviews, which were conducted with 5 dealers in April and May 2015. As I worked in the Project Management Office of the central project team for two years (08/2012 – 10/2014), I have profound knowledge of the project and access to the central and market implementation managers of the change initiative, who helped when choosing appropriate dealers for the interview. Dealers from the sales region “Northern Europe” were chosen for the case study as the heterogeneous circumstances of dealers operating in different markets ensure environmental variance (Pettigrew, 1990). Further, the implementation of the change initiative is completed at most of the dealerships, which is a requirement for the interview. The sales region “Northern Europe” contains 7 markets, namely Denmark, Finland, Norway, Sweden, Estonia, Lithuania, and Latvia. The total number of dealerships in Northern Europe adds up to 128 (Anon.). Each of the markets has a National Sales Company located in their country, serving as the main contact point for the dealers. The Regional Office of “Northern Europe” is in Sweden and serves as an interface to the Headquarters. Due to the relatively small market sizes, some of the departments, such as Dealer Development or Product Management, are centralized in the

Regional Office leading to the dealers contacting both their National Sales Company and the Regional Office, depending on the topic raised.

The 5 dealers interviewed operate in Denmark (2), Sweden (1) and Finland (2), which represent 3 out of the 4 biggest markets in “Northern Europe” regarding sales volume and number of dealerships. In Denmark, 2 dealers are operating a total of 10 dealerships; there are 49 dealerships in Sweden; and the Finnish market contains 20 dealerships (Anon.). The requirements for the dealers’ participation in the interview are a completed implementation of the change initiative, a minimum annual sales of 1,000 cars in all dealerships operated by the dealer, guaranteeing sales relevance for Northern Europe, and of course, a willingness of the dealer to participate in the interview. The interviewees were either board members of the dealer or sales managers, depending on the person in charge of the implementation of the change initiative at the dealerships. Within the interview session, one or two managers participated in the interview, depending on their field of responsibility, ensuring that all questions could be answered. As the focus of the case study is on the sensemaking of the dealer, interviewees work only on the management level, and thus there are no further interviews conducted with other employees at the dealership. Five interviews in total ensure that general conclusions about dealers operating in different market circumstances can be drawn, even within the limited time for this thesis.

### 3.3. Crafting instruments and protocols

The next step is about the data collection method. Theory-building studies usually use multiple data collection methods (Eisenhardt, 1989). The rationale is, similar to theory-testing, to obtain stronger confirmation of constructs. Data can be collected through diverse sources of evidence such as documentation, archival records, interviews, direct observations, participant observations or physical artifacts (Yin, 2003, p. 86). The so-called “triangulated methodology” builds on three different sources of evidence to draw on the particular and different strengths of various data collection methods (Pettigrew, 1990).

In line with former studies (Weber et al., 2015), the main source of evidence for this thesis will be in-depth interviews. The interviews are semi-structured and last approximately sixty minutes. Interview questions are derived from a case study protocol and are adjusted during the data collection process as a consequence of the emergence of new and interesting themes (Yin, 2003, p. 90; Eisenhardt, 1989). Each interview is taped and transcribed to ensure the accuracy of the followed analysis.

The biggest advantage of conducting interviews to gather data is that the researcher can focus directly on the case study topic and provide perceived causal inference. Possible disadvantages of interviews such as a response bias or a bias due to constructed questions, led to the decision to integrate further data sources. Archival documents are a good supplement as they can be reviewed repeatedly and are not created as a result of the case study. In order to get a better

understanding of the implementation process of the change initiative, archival documents of change and communication campaigns are considered (Yin, 2003, pp. 86).

To triangulate data, on-site observations were written down during the dealer visits. These observations contain both participants as well as site observations and are integrated into the field study notes (Guest et al., 2013, p.93).

In preparation of the data collection and to increase the reliability of the case study, a protocol is developed in accordance with Yin (2003). Using a protocol ensures that the researcher stays focused on the purpose of the case study and forces him to anticipate potential problems such as the way the case study reports are to be completed. The case study protocol contains an overview of the case study project, the field procedures, the case study questions as well as a guide for the case study report (Yin, 2003, p. 69). Thus, in contrast to a survey questionnaire, “the protocol contains the instrument as well as the procedures and general rules to be followed in using the protocol” (Yin, 2003, p. 68). The case study protocol for this thesis can be found in the appendix.

### 3.4. Entering the field

The next step of the research process determines the entering of the field. As briefly mentioned in the context of grounded theory, Glaser and Strauss (1967) recommend a joint collection and analysis of data. While the degree of overlap intended by Glaser and Strauss cannot be achieved in the thesis, the analysis of data will already take place during the data collection. As recommended by Eisenhardt (1989), field notes, which are a running commentary to oneself, can accomplish this overlap. Field notes serve as an idea booklet to document everything that seems important to the researcher.

Although the cases will be pre-defined, flexible and opportunistic methods allow for adding questions or even adapting the data collection methods. These adaptations are not to be seen as unsystematic, but as a controlled opportunity for the researcher to take advantage of the uniqueness of cases – space can be opened up for emergent concepts (Pettigrew, 1990).

### 3.5. Data analysis

The data analysis is decisive for the results of the case study. There are traditionally two forms of analysis. (1) If an analyst wants to convert qualitative data into a quantifiable form to test hypotheses, he codes the data first and then analyzes it. “A code is an abbreviation or symbol applied to a segment of words, in order to classify the words. Codes are categories that usually derive from research questions, hypotheses, key concepts, or important themes” (Miles and Huberman, 1984, p. 56). (2) When the analyst wants to generate theoretical ideas, he constantly redesigns and reintegrates categories and their properties (Glaser and Strauss, 1967, p. 101). This thesis follows the constant comparative method of qualitative analysis developed by Glaser and Strauss, which is a combination of the two forms of analysis.

When analyzing data, there are two different approaches of starting the coding. On the one hand, Miles and Huberman (1984) recommend creating a “start list” of codes prior to fieldwork. That list is derived from the theoretical framework, research questions and hypothesis (Miles and Huberman, 1984, p. 57). Glaser and Strauss (1967) on the other hand do not want to precode until data is collected, so data gets well molded to the codes that represent them and the analyst is more context sensitive. In accordance with Glaser and Strauss (1967) the data for this thesis is coded during the fieldwork.

The data analysis of this thesis starts with a within-case analysis to become familiar with each case as a stand-alone entity (Eisenhardt, 1989). Unique patterns can evolve, before being generalized through comparison. Based on detailed case study write-ups for each site, each incident will be coded into as many categories as possible so categories that emerge fit the data basis. Categories are either self-constructed (explanative), or abstracted (descriptive) from the language of the situation.

The next stage of the data analysis is the cross-case analysis. As researchers are overwhelmed with data and simultaneously limited in their ability to process this data, the key to a good cross-case comparison becomes looking at the data in many divergent ways (Eisenhardt, 1989). The approach of this thesis is to select categories first and then to look for patterns – either within-group similarities or intergroup differences. A pattern can further be characterized by frequency (things happen often or rarely), correspondence (things happen in relation to other activities or events), and causation (one causes another). In order to keep track of the comparison group, Glaser and Strauss define a rule for their constant comparative model: “while coding an incident for a category, compare it with the previous incident in the same category and look for different groups coded in the same category” (Glaser and Strauss, 1967, p. 106). The constant comparison of the incidents generates characteristics of the category. Underlying uniformities or properties within the categories will be discovered, so that categories can be consolidated. Thus, as the theory evolves, categories will be reduced. Over time, the researcher becomes committed to an evolving theory, so the original list of categories is reduced to collect and code data according to the boundaries of the theory. Then he can focus on the comparison of incidents in the reduced set of categories. As patterns between the categories evolve, evidence for “why” behind relationships is researched and insights are gained. The coding system of this thesis can be found in the appendix.

Figure 6 shows how a general theory is to be derived from a particular code.

### 3.6. Reaching closure

Eisenhardt (1989) names the last step of the research process “reaching closure” and embraces two issues: when to stop collecting more data and when to stop iterating between data and theory. “A researcher can always try to collect more data for checking hypotheses or for generating new

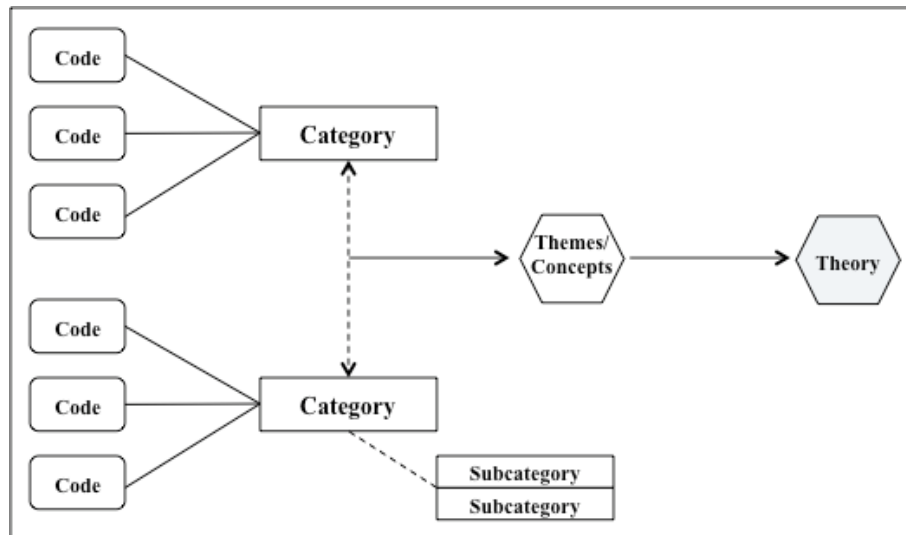
properties, categories and hypotheses” (Glaser and Strauss, 1967, p. 111). According to grounded theory, a researcher should stop adding cases when theoretical saturation is accomplished. This represents the point, at which additional learning is marginal because the researcher observes phenomena he has already seen before. The concept of saturation does also apply for the iterating process. The iteration process is completed when the incremental improvement to the theory is minimal. The application of theoretical saturation in this thesis is limited due to the predetermined amount of time.

After the theory-based research process for this thesis was outlined, provisions for the overall quality of the case study need to be met. To ensure the quality of the research study, Yin (2003) mentions four tests that are common to social research methods. “Construct validity is about establishing a correct operational measure for the concept being studied. . . Internal validity ensures a causal relationship, whereby certain conditions are shown to lead to other conditions. . . External validity identifies the extent to which findings can be generalized. . . Finally, reliability demonstrates that the operations of a study can be repeated with the same results” (Yin, 2003, p. 34). In Figure 7 the case study tactics, which have been discussed in the separate process steps, are related to the so-called four design tests.

## 4. Results

The final product of the theory building differs among concepts, conceptual frameworks and propositions. In line with former studies on sensemaking of middle managers (Balogun and Johnson, 2004; Balogun and Johnson, 2005; Hope, 2010), this thesis will report the findings of the data analysis using first- and second-order analysis according to Van Maanen (1979). The first-order analysis tells the story of the project implementation from the interviewees’ perspective and provides evidence for the coded themes through the use of representative quotes from the interviews. Both the descriptive properties of the studied scene and the interviewees’ interpretation of what stands behind these properties are first-order concepts. It is necessary to identify the change interpretations of the dealers, any other events they perceive to have an impact, and the change outcomes they observed (Van Maanen, 1979).

The second-order analysis explores the patterning of the first-order data and builds the explanatory framework. Many second-order concepts are statements about relationships between properties (Miles and Huberman, 1984, p. 104). Van Maanen (1979) calls second-order findings “interpretations of interpretations” as they represent the researchers’ perception of the interviewees’ described incidents. Therefore, second-order findings may converge with first-order interpretations. Hence, it should be understood how the dealer assigned meaning to the change interventions and how those perceived interpretations led to subsequent change outcomes (Van Maanen, 1979).



**Figure 6:** Codes to theory model; Source: Own figure, in accordance with Miles and Huberman, 1984, p. 57

| Tests                     | Case Study Tactics                                          | Phase of research in which tactic occurs |
|---------------------------|-------------------------------------------------------------|------------------------------------------|
| <b>Construct validity</b> | Multiple sources of evidence<br>Establish chain of evidence | Data collection<br>Research design       |
| <b>Internal validity</b>  | Pattern-matching<br>Explanation-building                    | Data analysis<br>Data analysis           |
| <b>External validity</b>  | Replication logic                                           | Research design                          |
| <b>Reliability</b>        | Case study protocol<br>Case study database                  | Data collection<br>Data collection       |

**Figure 7:** Case study tactics; Source: Own figure in accordance with Yin, 2003, p. 34

The bases for the first- and second-order analysis are within- and cross-case analyses, whose use has been previously described and which can be found in the appendix.

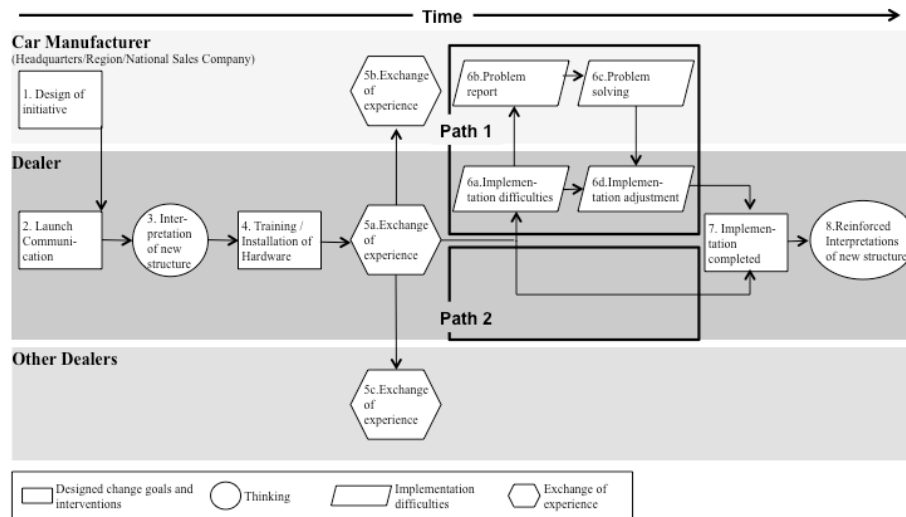
#### 4.1. First-order findings

The manually-obtained codes for change interventions, other events and outcomes based on the dealer's language can be found in the appendix. The detailed codes were gradually grouped into a set of broader categories related to:

- Evaluation of change initiative: first contact with the project, implementation progress, evaluation of the changes, project opportunities and challenges
- Evaluation of Retail Standards: The evaluation of Retail Standards and the dealers' ability to influence them
- Discursive practices within the dealer: Communication within dealer including management and employees, expectations of employees
- Discursive practices with other dealers: The overall relation and the communication with other dealers during the project implementation

- Discursive practices with the National Sales Company: The overall relation and the communication with the National Sales Company during the implementation, the evaluation of skills
- Discursive practices with the Region: The overall relation and the communication with the Regional Office during the implementation, the evaluation of skills
- Discursive practices with the Headquarter: The overall relation and the communication with the Headquarter during the implementation, the evaluation of skills

During the interview process, it became clear that the dealer's complex patterns of interaction during the change process, his implementation experiences and the organizational context in which he operates shaped his sensemaking. In accordance with Balogun (2003) a 'visual mapping strategy' is used to illustrate the perceived interactions (Langley, 1999). This serves as a basis for the first-order analysis that follows. The first-order analysis is divided into three parts (T1, T2, T3) based on the three steps of the sensemaking process, which were outlined in the literature review.



**Figure 8:** Visual mapping strategy; Source: Own figure

#### 4.1.1. Development of new structures at dealerships (T1)

This phase includes the design of the change initiative, the introduction to the dealers and his first interpretations.

The design of the initiative (1) took place within a project team at the Headquarters. The communication material includes an implementation plan providing information about the key milestones of the change initiative, the aimed timing and the distribution of responsibilities. Either departments within the Headquarters or the market, or an external implementation agency was in charge of the implementation process. The dealers didn't feel that they could influence the design of the change initiative or the implementation process:

'I think they come to us and that is how it is done. I don't have the feeling that I can impact on those.'

'We do not have the idea that we have an influence on anything. We got the idea that this is what the concept is about from the Headquarter.'

The launch communication (2) at the dealers took place differently. The communication material provided by the central project team contains a checklist for the National Sales Company's staff, which includes, "conduct dealer briefings to inform them about next steps". However, how the introduction of the change initiative to the dealer should take place is not predefined and thus is left up to the Region or National Sales Company, which mainly interacts with the dealer. For that purpose, the Region and National Sales Companies had a variety of information material from the Headquarters, such as the "EPOS Guide and Furniture Argumenter", which is to "be forwarded to the dealer".

Region representatives introduced the change initiative to the Danish and Swedish dealers. In Finland, a consultant presented the concepts to the dealers. Although the launch communication was different, the dealers commonly spoke

of an abstract presentation, lacking concrete consequences for them at that time.

'... And they presented this like they presented it in Germany or in Poland or somewhere else.'

'I think he didn't know himself, too when he was presenting the ideas for the first time.'

'... It was not so concrete. I didn't actually know what was going to happen. It was sort of a nice picture somewhere but no concrete footsteps like 'now we are doing this, this, and this and the schedule is this. So, it was a little bit abstract at that time.'

'It is a little bit more the big picture, but not that concrete.'

Although, the first picture was abstract, the dealers commonly thought that the sales strategy led to significant changes at their dealership including the standardization of sales processes, an enhanced customer experience, the modernization of sales tools and the showroom appearance. The interpretations of the new structure (3) were ambiguous, perceiving the overall ideas as positive, but the implementation as difficult:

'We strongly believe that this could be a success.'

'I think it is a big step to take. But it is the right step to take.'

'I think that was great that we did something. Because I think that it was quite old fashioned in (franchisor).'

'On paper the ideas are very good, but then how to use everything in the daily work, that is our challenge at the moment.'

'I think they have not thought it through, how to take this step and how to change the whole processes.'

'It seemed SO much at the same time, at once. I



think the ideas are good. But I think they are trying to implement it too fast and have not really thought it through.'

#### 4.1.2. Implementation of new structures at dealerships (T2)

This phase includes the training and installation of the previously described concepts; the communication within the dealer, with other dealers and with the franchisor; as well as the handling of implementation difficulties.

The implementation of the change initiative is for the most part completed at the visited dealerships. At the Danish and Swedish dealers, the new furniture, the mobile tools, the Product Genius and Mystery Shopping were implemented. Due to the smaller size of the dealerships, the interviewed Finnish dealers had not yet implemented the Product Genius.

During the installation and training (4), the interviewees gained information on the new sales strategy differently. Either they were trained themselves, or gathered information through their employees who attended a training session. For some, manuals served as an additional information source, whereas others did not use them or did not have any. Overall, some dealers said that they received enough and detailed information from the market whereas others claimed they had to figure out how to implement the concepts themselves:

'The trainings have been helpful.'

'We had a lot of persons that came and told us about the implementation. How it is going to work and why of course we have to do this. So that was good.'

'We had one training in here just a few weeks ago, also again with the consultant (name). He showed us the use of the iPad, so now everyone should know how to use it.'

'The Product Genius had the training and we have heard what they are training.'

'We have very big manuals. And then we have a lot of follow-ups'

'We derived our knowledge from the manuals, from the Retail Standards and from the test answers from the Mystery Shopper.'

'We don't have any instruction manual.'

'Yes, we make it by ourselves. Nobody answers to that. And when we implement the Product Genius for instance or set up the Mobile Customizer, we had to do everything ourselves. Not any support on that.'

Besides the training and information material, the interviewees gained expertise through an exchange of experience (5) with different people. Most of the interaction took place at the dealer (5a). Some dealers conducted their own trainings in addition to those provided by the franchisor in order to fulfill the Retail Standards. Further, they had regular meetings at both the management level and with the sales staff, where they exchanged experiences:

'The rules are set from (franchisor) but the daily training is in the shop.'

'We know that we have a how-to-do-it from (franchisor) but everyday we are reminding ourselves, training, so we are making sure that we give every customer the right experience when they come to the (brand) stores.'

'We do it when we train and when we have our meetings. What is good, what needs to be improved. That is mainly in-house we do that.'

'In our house, we have talked about it daily, how to do things better.'

'We have a small team here. We often talk directly how we can improve things; how we can do things better. Also, I like to listen to my sales guys, what their opinions are, how they see problems or issues.'

For further clarification, the dealer saw the interaction with the National Sales Companies and Region (5b) as important. The relationship with the National Sales Companies and the Region was described as very positive. However, while some dealers perceived the interaction with the franchisor as an exchange of experience, others perceived it as a unidirectional communication.

'The cooperation is very good.'

'Our Region Manager is very, very good. I think he is one of the best that I have ever met. And one of the best I have ever worked with at (franchisor).'

'Yes, I talk to him on the phone every week and then we meet at least one time a month.'

'It is always a one-sided communication.'

'Since I worked for (franchisor) I did not have the impression that they like our comments on anything and try to receive feedback.'

All interviewees said, that they exchanged experience with other dealers (5c) informally. They did not schedule meetings nor interact on a regular basis:

'When we are meeting for sales meetings, or something else, then we just chat to each other and have a conversation. But it is nothing that is put into meetings or schedules.'

During this phase, some of the interviewed dealers faced implementation difficulties, whereas for some the implementation went very smoothly. Whether or not the dealers had implementation difficulties, partially influenced the change outcomes and their interaction processes.

For the dealers, which faced implementation difficulties, the type of problem and the solution finding were considered in more detail during the interviews. The implementation difficulties (6a) mainly routed in the employees' acceptance of the new, more standardized sales process and infrastructure problems.

'I know the salesmen are afraid. They are afraid to do wrong. This becomes so much of a burden. So the only thing they are thinking about is not to do wrong.'

'Before this, they have always been doing it in the same way and people get stuck to their own ways of doing things. That is the problem.'

'I think the main problem about the Product Genius is their ambitions to be a salesman.'

'We lack of systems, sales systems actually, to help the sales men in all that process.'

'The VPP does not work so they cannot use the iPad at the moment.'

'Today we have to go into 5-6 different computer programs just to make an offer...When (franchisor) solves that problem everything will fall into place.'

'First, we built it with the iPad; as we cannot link it to SA3, we have to do the same thing within SA3 again; and then, when we have done it there, we will do it within the CRM system.'

When a problem occurred, the dealers first tried to solve it on their own, before reporting it to the National Sales Company or Region (6b).

When involving the National Sales Company or Region, the main problem was the lacking ability of the franchisor in solving the problems (6c). When an implementation problem occurred, most of the dealers stated that the National Sales Company or the Region were not able to solve it.

'I think the support should be at the importer. But now, if I had any kind of problem I can't call them. No one knows about it. They haven't used it.'

'I think that neither the consultant nor the importer know how to use it. So we have to figure it out ourselves.'

'Yes, actually it is quite amusing that they tell us that we have to use it but if we have some questions, they can't answer them.'

'They just did not do their homework how to solve the problems that come along.'

'It has gone almost two years and we are still waiting for them to come and solve the problem'

'We are waiting that something will happen.'

'The sales guys should already start in 2009 configuring the cars with the customers on a big screen and so on. But it has been postponed for six years.'

The interviewees identified the lack of retail experience as the main source for the National Sales Company's or Region's lacking ability to find solutions for their implementation problems. In addition, the project didn't fit their market circumstances completely.

'I think importers and factories have one big problem or challenge: the challenge is, they don't sell cars. They don't see the customer, they don't know how to treat a customer, and they don't know what the customer expects.'

'No manager or importer, have sold a car before. They say, what they have to say but they don't really understand how the work is on the floor. They don't have retail experience.' 'They know a lot about wholesale, but nothing about retail.'

'Most of the employees at (franchisor) have never been in a dealership.'

'And I am not sure if the importer knows what he is talking about. Someone told them that (change initiative) has to be implemented and they say that to us.'

Furthermore, the implementation plan of the Headquarters does not contain steps that the National Sales Company or the Region should follow, when implementation difficulties occur.

For the problems that were solved, either in-house or with the help of the National Sales Company, the implementation was adjusted (6d).

#### 4.1.3. Completion of implementation and evaluation of new structures (T3)

This phase includes the completion of the implementation and the dealers' evaluation of the change initiative.

The implementation of the change initiative at the interviewed dealers is for the most part completed (8). Through the dealer visits it could be observed that the appearance of the showrooms was indeed standardized. The elements of the sales strategy were remarkable in all showrooms. Although the size of the visited dealerships differed, the design of the workplaces and the waiting areas for the customers looked the same.

After the outcomes of the initiatives were observable in the adapted sales processes, the dealer still perceived the overall targets as positive. However, when looking at the individual concepts, the dealers partially rated the outcomes differently (7).

The Mobile Customizer, which has been implemented at all visited dealerships, was rated similarly. It was mainly perceived as a support and modernization of the sales process. Yet, most of the dealers criticized the app for not being harmonized with existing IT systems, leading to inefficiencies in the sales process.

'It is a very good idea. I think it is much better than what we had before.'

'I think these are good tools, which can support the selling process.'

'We have something, at the moment, if a customer is using our (franchisor) website, the configuration tool, we cannot implement that to SA3

directly or to the Mobile Customizer. We have to build it again.'

'And we hope that this system can one day support SA3. Because that would be the best way for us, we would get better advantages if we could configure a car with the iPad and then just send the information to SA3.'

In the case of the new furniture, the interviewees valued the consistent, premium showroom appearance and the high acceptance of the employees. However, the high investments were perceived as a risk.

'The employees of course they liked it... It is always nice to get new furniture and it upgrades the feeling of course.'

'I think it is nice that the showrooms look alike no matter where you go'

'It is difficult to get the business case profitable, with the total costs of this showroom, of all the standards of (franchisor).'

'It has a big influence on our economy.'

As previously mentioned, not all of the interviewed dealers have employed a Product Genius at their dealerships. This is mainly due to the different sizes of the dealerships. It is more difficult for a dealership with fewer customers to have a full-time employee who is used to the capacity of the tasks of a Product Genius. Thus, implementing the concept would not be profitable. The perception of the dealers who have employed a Product Genius was similar. On the one hand, the new role led to an enhanced customer service through the Product Genius' product expertise and his customers' need-analysis. Further, the work division between the Product Genius and the salesmen led to higher efficiencies in the sales process. On the other hand, the dealers had difficulties managing the expectations of the Product Genius, who often had ambitions taking over tasks of the salesmen.

When looking at the dealers' evaluation of Mystery Shopping, it is important to consider that the results of the Mystery Shops are not part of the Dealer Bonus in all three markets. This could be one explanation for the different ratings of Mystery Shopping. While some dealers perceived it as a helpful tool to increase customer satisfaction and gain transparency over the sales process, others said that it puts too much pressure on the employees, affecting the dealers' results negatively.

'The report is helpful because it gives us an insight of how customers think when they are coming here. It also gives us an insight where we could improve, which salesmen are doing a good job and which we need to train more.'

'I think it is a good thing to keep us sharp and remember that we have to do it right and that we have to do it right each time.'

'If we fail, it cannot be that we are going to get so hardly punished.'

'It damages the treatment of real customers. The salesmen are always thinking 'Are you a Mystery Shopper?' They are always suspicious, afraid to do wrong.'

#### 4.2. Second-order findings

Included in the first-order analysis are descriptive findings from the interviewees' perspective about how and why different change outcomes emerged. In order to find theoretical explanations for these descriptive first-order findings, a second-order analysis follows.

A key factor in the dealers' sensemaking is the social processes of interaction in response to the design and launch of the new structures (T1), the installation of the change initiative at the dealer (T2) and the evaluation after the completed implementation (T3). The first-order analysis reveals two dimensions of social processes of interaction, shaping the dealers' sensemaking, which serve as a basis for the explanatory framework: the social processes of interaction within these phases can be divided into vertical processes, which go beyond organizational boundaries and take place between the dealer and the car manufacturer or between the dealer and other dealers; and lateral processes, which stay within the dealer. It can be seen that both vertical and lateral processes of interaction affected the sensemaking of the dealer, but at different stages.

In T1, the dealers' sensemaking was triggered by the change intervention itself and the initial communication by the franchisor. The change initiative constituted a novel and unexpected event for the dealer, which had a big impact on his organizational identity. As the communication of the change initiative implied ambiguous interpretations, the dealer involved himself in a sensemaking process. In this phase, mainly vertical processes of interaction took place. Employees of the National Sales Company set meetings with the dealer and introduced the change initiative themselves or sent a consultant to present the main targets to the dealer. The dealers commonly identified the premium customer experience and the standardization of sales processes as the cornerstones of the visions. As these cornerstones are commonly perceived as reasonable, the overall idea was well understood and accepted. However, the new vision of the change initiative was not commonly created, but rather expressed through the National Sales Company, which can be seen as sensegiving. The dealers would have preferred a common sensemaking, so their insights were considered during the design of the change initiative. Although the first meaning creation by the dealer assigned positive attributes to the vision of the change initiative, the feasibility of applying them to his own environment was seen critical. The dealer found that he had not sufficient information through the first presentation. Thus, he was trying to figure out the meaning of the change effort, what its effect on him would be and what his role would entail. By the end of T1, the dealer had the awareness that a new way of thinking about

retail structures were to come, but he was lacking knowledge about concrete actions to follow.

In T2, the implementation of the new structures at the dealer began. The installation of new hardware and the training sessions at the dealer were organized by the franchisor. Thus, the social processes of interaction were mostly vertical at first. During the training sessions, the dealer could now gather more concrete information through discussions with representatives of the franchisor and partially obtained additional information material on the individual concepts. The dealer's perception of the franchisor's support during that phase differed. Although the relation to the franchisor and his intentions are positively perceived, his ability to convey sufficient information to the dealer is questioned.

As there was still information lacking, the dealer exchanged experience, both laterally and vertically, after the official training sessions given by the franchisor. Here, dealers mostly conducted their own trainings or engaged in internal discussions to clarify open issues. So, the dealer involved his employees into a common sensemaking process to integrate their expertise on the dealer's circumstances in order to create a common understanding on the implications of the change initiative. The input of both the employees and the management at the dealer, as well as the processes of social interaction between them were valued. Through the involvement of the employees in the sensemaking process, legitimacy was created resulting in the employees' acceptance of the changes in their working routines.

The vertical interaction processes with other dealers did not take place on a regular basis and did not seem to play a decisive role in the dealers' sensemaking.

The dealer provided the franchisor with feedback at that point but for the most part, no common sensemaking activities followed. The dealer did not think that his feedback led to adaptations of the initiative or that the National Sales Company forwarded the feedback to the Headquarters, where the change initiative had been designed.

During the implementation, some dealers faced difficulties. The main challenges were the adaptation of the employees' working routines and the occurrence of infrastructure problems such as the harmonization of the new tools with existing systems. Dealers complained that the implementation of the concepts at the dealership was not thought through in detail by the franchisor. Mainly concepts were not harmonized with existing IT systems or Retail Standards in advance. These struggles created further ambiguity and led to additional sensemaking activities at the dealer. In most cases, the dealer tried to solve these problems internally. He provided his employees with additional trainings and collected their feedback in regular meetings to help integrate the new Retail Standards within daily work. Furthermore, the dealer involved the franchisor asking either for additional training sessions or for technical support. However, the interaction with the franchisor mostly did not lead to a solution. The dealer often got the impression that the franchisor could not provide support during the integration of the change initiative in his operative business as the franchisor lacked retail

experience and did not have enough knowledge on the detailed functionalities of the concepts.

In T3, the implementation of the change initiative at the dealer was completed and outcomes could be observed, which served as input for new sensemaking. At this point, the dealer mostly observed positive impacts on the results of his business unit such as increased customer satisfaction and sales volume. The dealer was able to learn about the change by taking actions and observing the cues generated by his actions. Regular meetings at the dealership secured a continuous exchange of experience between the employees, who applied elements of the change initiative in the sales process. Mystery Shopping at the dealerships provided input for both the franchisor and the dealer, who could review if the elements of the change initiative had been integrated into the sales process.

Further, the application of the change initiative at the dealership tested the dealers' provisional understanding generated through his prior sensemaking activities. Lastly, the environment at the dealership has been changed through the implementation of the change initiative and therefore changed the situation that initially triggered the dealer's sensemaking.

## 5. Discussion

The idea that social processes of interaction impact the sensemaking activities of organizational actors during change is well accepted (Balogun and Johnson, 2004; Balogun and Johnson, 2005; Hope, 2010). The contribution of the framework presented in this thesis is the distinction between vertical and lateral processes of interaction and the empirically-based concept it offers to explain how sensemaking takes place during inter-organizational change. This thesis links inter-recipient sensemaking activities within and across organizational boundaries to implementation outcomes. Key is not that social processes of interaction during sensemaking occur, but how they influence individuals' interpretations when the change design and the change implementation take place at different organizations.

Generally speaking, the findings show that a franchisee, who operates his own business and has his unit results at stake, has the pressure of implementing the change initiative well. Thus, he has an increased pressure of involving himself in sensemaking activities.

The second-order analysis reveals that organizational boundaries hinder common sensemaking activities between the franchisee and the franchisor. Franchisors provide the franchisee with an already-established vision and implementation steps to follow. A lack of prior common sensemaking activities during the change design leads to the franchisee's lack of trust and belief in the legitimacy of the change initiative. As other studies on the role of franchisees during change suggest, trust in the expected profitability of the change initiative has a high impact on the adaption of a constructive response by the franchisee. Further, it is shown that the involvement of franchisees in the strategic decision-making

process can create this trust (Croonen, 2010). Looking at studies on sensemaking, it is also shown that the involvement of stakeholders in the design of the change initiative creates legitimacy (Gioia and Chittipeddi, 1991).

When formulating a new vision, the franchisor shall integrate unquestionable goals, which serve his and the franchisees interest at the same time (e.g. higher customer satisfaction, higher sales volume). Similar to other studies on sensemaking during change, the case study of this thesis shows that these unquestionable goals lead to a shared understanding and a positive acceptance of the overall vision (Gioia and Chittipeddi, 1991).

However, when planning the implementation steps at the franchisees' business unit, the franchisees' trust in the franchisors' ability to improve his business processes is low. Although the franchisor and the franchisee share overall goals, the franchisees' business interest is based on operative experience whereas the franchisor is thinking more strategically, having the overall targets of the franchisor in mind. These different perspectives and also past experiences evolving from different working routines, lead to different interpretations of the change design. Thus, involving in common sensemaking activities would help to share these perspectives and create a common picture.

Beyond that, studies on franchising reveal that trust in the franchisor influences the response of franchisees during change (?). This thesis shows that franchisees' trust is not only influenced by the franchisors' operative skills, but is also affected through his troubleshooting during the change implementation.

Similar to the studies on the role of middle managers during change (Balogun and Johnson, 2004), most of the franchisees' sensemaking activities during the change implementation occur in the absence of the actor, who designed the change. The most important sparing partners of middle managers are on the same hierarchical level and entail a similar role during the implementation (Balogun and Johnson, 2004). Transferring that to franchisees, the exchange of experience with other franchisees is rather informal and has little impact on the sensemaking of franchisees.

The franchisee mainly integrates the employees of his business unit into his sensemaking activities. However, in contrast to middle managers, who share their understandings in unofficial settings such as rumors, stories or gossip, the franchisee organizes official meetings and trainings to create common interpretations. Thereby, experiences of applying the change initiative into daily working routines are shared among managers and employees. The creation of shared interpretations also creates legitimacy among the employees towards the change design and reduces their uncertainties towards their new working routines.

Besides the interaction processes during the implementation, this thesis shows how franchisees further involve in sensemaking through the observation of change outcomes at their business units.

### 5.1. Practical implications

The franchisor provides a plan for change, but the way this plan actually operates is determined by the new behavioral routines created by the franchisees through their interpretation of and response to franchisors' initiatives. Franchisees can react differently to franchisor-initiated changes. As destructive responses can seriously harm the franchise system and thus the retail performance of a company, the franchisor should anticipate and prevent the franchisees' destructive reaction (?). Analyzing the franchisees' sensemaking activities provides useful insights for the franchisor planning to implement change initiatives at his franchisees.

Since the findings show the key role franchisees entail during franchisor-initiated change and how they accomplish sense of a planned initiative, like other studies, this thesis questions the feasibility of top-down control of change programs (Balogun and Johnson, 2005). Top-down control in inter-organizational change refers to the control the franchisor exerts on the franchisee.

Managing change includes the creation of trust and legitimacy. Thus, looking at inter-organizational change from a sensemaking perspective shows that managing change is less about directing and more about supporting franchisees' sensemaking processes to achieve an alignment of interpretations. This is especially important in franchising, where the franchisee carries more economic risks than actors within organizations. Shared sensemaking activities and aligned interpretations, in both the change design and the implementation, create positive acceptance of the franchisee towards the change initiative and helps the franchisor to gain insights into the operative business processes. Involving the franchisee in the design of new structures in their market and adapting the initiative together to market circumstances, increases legitimacy at the franchisee. Further, an increased knowledge and consideration of the business processes at the franchisee, creates trust in and trustworthiness of the franchisor, which constitutes a critical factor of success. The franchisor is not able to manage lateral processes occurring in his absence, but he can shape his interactions with the franchisee. Besides considering the franchisees' insights in the change design, trust is also accomplished through efficient troubleshooting in the change implementation.

### 5.2. Limitations

The limitations of this thesis need to be considered. Questions of generalizability arise, as case study research is, by nature, situational. This is particularly influenced through the focus on franchising systems, which is an extreme case of inter-organizational change. The research approach, the case study setting, and the findings are interwoven.

Further, although, environmental variations were considered in the case study sampling, the number of interviewed franchisees is limited due to the time restrictions of this thesis. As the interviews served as the primary data source, the actions taken by the franchisor and the reactions of the employees at the franchisees' business unit are mainly antici-



pated based on the franchisees' responses. Thus, the display of the change process can be biased.

Although three different data sources were considered for the analysis, the researcher was not able to observe the interviewees' real-time behavior during the change process. Lastly, the researcher's tie to the franchisor, through prior work experience, might have influenced the responses of the franchisees.

Despite these limitations, this thesis provides a useful foundation for future studies on sensemaking during inter-organizational change.

### 5.3. Future research

One promising direction for future research could be the investigation of the complete sensemaking-sensegiving process during inter-organizational change. Thereby, the sensegiving activities across organizational boundaries, initiated by both the franchisor and the franchisee, could be of interest. In addition, analyzing the sensemaking-sensegiving activities at the franchisee, considering the management and the employees, could create useful insights. The findings of this thesis suggest that the adaptation of the franchisees' employees' behavior has a high impact on the success of change initiatives. Thus, the adaptation process of their working routines should be investigated. Interaction processes with their management and, across organizational boundaries, with the customer might influence their sensemaking activities. In the end, the customers' reaction and buying behavior constitutes the success of the change initiative. Besides further investigating franchisor-initiated change at franchisees, this thesis could serve as a basis for investigating other forms of inter-organizational change, considering different industries or different stakeholders.

## 6. Conclusion

This thesis has examined the sensemaking of franchisees' during franchisor-initiated change with the use of a holistic multiple case study. It is the first to analyze sensemaking during inter-organizational change.

Overall, it was shown that franchisees play an important role in implementing strategic change. Their actions can either support or be convergent to change goals set by the franchisor. Thus, understanding the processes through which franchisees' work to assign meaning to a change initiative is of highest interest to the franchisor.

The analysis revealed that franchisees' social processes of interaction mainly affect their sensemaking. A distinction between lateral and vertical processes of interaction enabled a deeper analysis of processes within and across their organizational boundaries. So far only few authors analyzed the social process of interaction during sensemaking (Balogun and Johnson, 2004; Balogun and Johnson, 2005; Hope, 2010). In contrast to studies on intra-organizational change, it was shown that franchisees mainly interact within their organizational boundaries when forming change interpretations.

The communication across organizational boundaries with the franchisor was often unidirectional. Thus, not many common sensemaking activities took place, which created the risk of misaligned change interpretations of the franchisor and the franchisee. This finding led to the practical advice to franchisors to align their interpretations with the franchisees' through more common sensemaking activities during the change design and implementation. The involvement of franchisees creates trust and legitimacy, and thus it increases the chances of a positive change outcome.

In accordance with other studies (Gioia and Chittipeddi, 1991), the findings of this thesis show how the integration of shared and unquestionable goals into the vision formulation creates positive change interpretations at the franchisee.

The thesis also finds similarities to studies, which identify the observation of change outcomes as additional input for sensemaking activities (Maitlis and Christianson, 2014). The identification of underlying similarities in phenomena increased the internal validity and generalizability of this thesis, which is especially important in theory building as the number of cases is limited. The differences to previous studies underline the relevance of the study conducted.

Future research should aim to better understand the overall sensemaking-sensegiving process for all actors during inter-organizational change.

## References

- Balogun, J. From blaming the middle to harnessing its potential: Creating change intermediaries. *British Journal of Management*, 14(1):69–83, 2003.
- Balogun, J. and Johnson, G. Organizational restructuring and middle manager sensemaking. *Academy of Management Journal*, 47(4):523–549, 2004.
- Balogun, J. and Johnson, G. From intended strategies to unintended outcomes: The impact of change recipient sensemaking. *Organization Studies*, 26(11):1573–1601, 2005.
- Bartunek, J. M. Changing interpretive schemes and organizational restructuring: The example of a religious order. *Administrative Science Quarterly*, pages 355–372, 1984.
- Bartunek, J. M., Krim, R. M., Necochea, R., and Humphries, M. Sensemaking, sensegiving, and leadership in strategic organizational development. *Advances in Qualitative Organizational Research*, 2:37–71, 1999.
- Basu, K. and Palazzo, G. Corporate social responsibility: A process model of sensemaking. *Academy of Management Review*, 33(1):122–136, 2008.
- Burns, T. and Stalker, G. M. The management of innovation. London: Tavistock. 1961.
- Cameron, K. S. and Quinn, R. E. Organizational paradox and transformation. *Cambridge: Ballinger*, 1988.
- Charmaz, K. 'discovering' chronic illness: using grounded theory. *Social Science & Medicine*, 30(11):1161–1172, 1990.
- Clegg, S. R., da Cunha, J. V., and e Cunha, M. P. Management paradoxes: A relational view. *Human Relations*, 55(5):483–503, 2002.
- Cornelissen, J. P. Sensemaking under pressure: The influence of professional roles and social accountability on the creation of sense. *Organization Science*, 23(1):118–137, 2012.
- Cox, J. and Mason, C. Standardisation versus adaptation: geographical pressures to deviate from franchise formats. *The Service Industries Journal*, 27(8):1053–1072, 2007.
- Croonen, E. Trust and fairness during strategic change processes in franchise systems. *Journal of Business Ethics*, 95(2):191–209, 2010.
- Dada, O., Watson, A., and Kirby, D. A. Toward a model of franchisee entrepreneurship. *International Small Business Journal*, 30(5):559–583, 2012.
- Davies, M. A., Lassar, W., Manolis, C., Prince, M., and Winsor, R. D. A model of trust and compliance in franchise relationships. *Journal of Business Venturing*, 26(3):321–340, 2011.
- Eisenhardt, K. M. Building theories from case study research. *Academy of Management Review*, 14(4):532–550, 1989.
- Ford, J. D. and Ford, L. W. Logics of identity, contradiction, and attraction in change. *Academy of Management Review*, 19(4):756–785, 1994.
- Garfinkel, H. Studies in ethnomethodology, Englewood Cliffs NJ: Prentice Hall., 1967.
- Gioia, D. and Thomas, J. Institutional identity, image and issue interpretation. *Administrative Science Quarterly*, 41(3):370–403, 1996.
- Gioia, D. A. and Chittipeddi, K. Sensemaking and sensegiving in strategic change initiation. *Strategic Management Journal*, 12(6):433–448, 1991.
- Glaser, B. and Strauss, A. The discovery of grounded theory, London: Wiedenfeld and Nicholson, 1967.
- Guest, G., Namey, E., and Mitchell, M. Collecting qualitative data – a field manual for applied research, Beverly Hills CA: Sage Publications., 2013.
- Helms Mills, J., Thurlow, A., and Mills, A. J. Making sense of sensemaking: the critical sensemaking approach. *Qualitative Research in Organizations and Management: An International Journal*, 5(2):182–195, 2010.
- Hernes, T. and Maitlis, S. Process, sensemaking and organizing, Oxford: Oxford University Press., 2010.
- Hope, O. The politics of middle management sensemaking and sensegiving. *Journal of Change Management*, 10(2):195–215, 2010.
- Langley, A. Strategies for theorizing from process data. *Academy of Management Review*, 24(4):691–710, 1999.
- Lewis, M. W. Exploring paradox: Toward a more comprehensive guide. *Academy of Management Review*, 25(4):760–776, 2000.
- Louis, M. R. Surprise and sense making: What newcomers experience in entering unfamiliar organizational settings. *Administrative Science Quarterly*, pages 226–251, 1980.
- Lüscher, L. S. and Lewis, M. W. Organizational change and managerial sensemaking: Working through paradox. *Academy of Management Journal*, 51(2):221–240, 2008.
- Maitlis, S. The social processes of organizational sensemaking. *Academy of Management Journal*, 48(1):21–49, 2005.
- Maitlis, S. and Christianson, M. Sensemaking in organizations: Taking stock and moving forward. *Academy of Management Annals*, 8(1):57–125, 2014.
- Miles, M. and Huberman, A. M. Qualitative data analysis, Beverly Hills CA: Sage Publications, 1984.
- Milliken, F. J. Three types of perceived uncertainty about the environment: State, effect, and response uncertainty. *Academy of Management Review*, 12(1):133–143, 1987.
- Pettigrew, A. M. Longitudinal field research on change: Theory and practice. *Organization Science*, 1(3):267–292, 1990.
- Polanyi, M. Sense-giving and sense-reading. *Philosophy*, 42(162):301–325, 1967.
- Polkinghorne, D. E. Narrative knowing and the human science, Albany NY: State University of New York, 1988.
- Putnam, L. L. Contradictions and paradoxes in organizations. *Organization-Communication: Emerging Perspectives*, 1(151-167), 1986.
- Rouleau, L. and Balogun, J. Middle managers, strategic sensemaking, and discursive competence. *Journal of Management Studies*, 48(5):953–983, 2011.
- Salancik, G. R. et al. Commitment and the control of organizational behavior and belief. *New Directions in Organizational Behavior*, 1:54, 1977.
- Schutz, A. *The phenomenology of the social world*. Northwestern University Press, 1967.
- Sonenshein, S. We're changing—or are we? untangling the role of progressive, regressive, and stability narratives during strategic change implementation. *Academy of Management Journal*, 53(3):477–512, 2010.
- Van Maanen, J. The fact of fiction in organizational ethnography. *Administrative Science Quarterly*, 24(4):539–550, 1979.
- Vince, R. and Broussine, M. Paradox, defense and attachment: Accessing and working with emotions and relations underlying organizational change. *Organization Studies*, 17(1):1–21, 1996.
- Weber, M. S., Thomas, G. F., and Stephens, K. J. Organizational disruptions and triggers for divergent sensemaking. *International Journal of Business Communication*, 52(1):68–96, 2015.
- Weick, K. E. *Sensemaking in organizations*, volume 3. Sage, 1995.
- Westenholz, A. Paradoxical thinking and change in the frames of reference. *Organization Studies*, 14(1):37–58, 1993.
- Yin, R. Case study research (3rd edition), Beverly Hills: Sage Publications, 2003.



# The Influence of Cryptocurrencies on Enterprise Risk Management – An Empirical Evidence by the Example of Bitcoin

Maximilian Bölstler

Universität St. Gallen

## Abstract

This thesis analyzes the influence of cryptocurrencies in the context of risk management by considering the emerging risk factors of Bitcoin as a payment method. By means of an empirical analysis through an online survey, the current operational dealing of incoming Bitcoin funds, the risk awareness of the potential threats, and the corresponding control activities implemented by companies accepting Bitcoin payments have been examined. The results reveal that the risks of this new technology-based payment method have not been extensively evaluated and that there exists a partially significant lack of know-how. Therefore, the risks are either not at all or improperly addressed by a majority of the organizations. However, the exchange rate risk and the cyber risk, which is a strongly linked to the administration of cryptocurrencies, represent the most significant related risk factors associated with cryptocurrencies in recent times. To ensure an appropriate operational dealing with cryptocurrencies, the author presents a risk control matrix based on the results of the analysis and discusses control activities to mitigate these emerging threats. Finally, a holistic Cryptocurrency IC Framework (following the COSO 2013 IC Framework) is presented, with the objective of effectively and efficiently developing and maintaining systems of internal control with regard to cryptocurrencies.

**Keywords:** Blockchain; digital assets; Bitcoin; cryptocurrency; IC framework; enterprise risk management.

## 1. Introduction

### 1.1. Problem Statement & Research Questions

The 21st century is characterized by continuous changes and the need for ongoing adaptation, with key terms like "digitization" and "automatization" playing an increasingly important role, particularly in the fast paced environment of a company. These changes can in fact be observed in almost all industries, ranging from manufacturing businesses to service providers. In these modified circumstances, with the emergence of new opportunities, unfamiliar risks and threats have also arisen. It is thus inevitable for companies and their executive management teams to thoroughly think through their implemented business models and, moreover, to adapt to the changing conditions for continued growth (Rüegg-Stürm and Grand, 2015, p. 78).

The concept of blockchain emerged in 2008 as an innovative technology in the sphere of currencies and payment methods. Initiated by the pseudonymous Satoshi Nakamoto (2008, pp. 1f.), the blockchain was a new idea for an electronic cash system based on cryptography resulting in the so-called "cryptocurrencies." The concept aimed at transforming

how transactions were conducted, leading to fundamental changes in the financial system and influencing its structures behind. After the financial crisis of 2008, the related loss of confidence in the prevailing financial system, and recent discussions of the International Monetary Fund (IMF) about the elimination of hard cash (IMF, 2017, pp. 4f.; Papadopoulos, 2015, p. 160), the cryptocurrencies have caught the attention of innumerable individuals, companies, and entire nations.

There are several kinds of cryptocurrencies, and currently, the most famous and widespread is "Bitcoin." The extent and prevalence of the nine-year-old cryptocurrency can be demonstrated by its considerable market cap, which was approx. USD 148 bn. as of February 4, 2018 (Coinmarketcap, 2018, 20. January). According to recent statistics of the Central Intelligence Agency (2018), this is equivalent to the actual stock of narrow money (M1) of countries such as Venezuela and Finland.

In general, Bitcoin can be used as medium of exchange or as a storage of value, i.e., as an investment (de Jong, 2015, p. 416). Some retailers already accept Bitcoin as an additional

payment method to purchase their goods and services. The most frequently mentioned advantages of digital currencies as a payment method are the speed, the price savings, and the increased security of (global) transactions, without the necessity for neither a customer nor a vendor to be a registered bank client and, moreover, without the involvement of any financial institution in the transaction process (Bank for International Settlement [BIS], 2015, p. 3).

At first glance, this innovation seems impressively progressive, as it is a payment process that is more efficient, has almost no transaction fees, and has fewer intermediaries. However, the novelty of the technology also introduces uncertainty and unfamiliar risk factors for the participants of the system. Consequently, the enterprise risk management (ERM), particularly of companies and retailers accepting cryptocurrencies as a payment method, will face new challenges.

Due to a lack of research on the appropriate handling of cryptocurrencies in business operations, this thesis aims at developing a holistic framework that supports companies accepting cryptocurrencies in establishing an appropriate risk management system for dealing with cryptocurrencies. Therefore, the following research questions will be addressed:

- What are the potential risks and resulting threats of accepting cryptocurrencies as a payment method?
- Are the companies and the responsible employees aware of these risks, and how do they evaluate and address them?
- Are there any general control activities and recommendations necessary to ensure an appropriate risk management in the handling of cryptocurrencies?

## 1.2. Structure

This paper has been structured as follows: The introduction (Chapter 1) emphasizes the relevance of the topic, the lack of research, and presents the research questions that will be addressed in order to examine the influence of cryptocurrencies on risk management by the example of Bitcoin. Subsequently, Chapter 2 presents an intensive literature review on the topic of cryptocurrencies and ERM. Chapter 3 presents the research approach adopted, which clarifies the research concept and the methodology employed. On this basis, the results containing the empirical investigation follow in Chapter 4, which is divided into three linked sub-sections, one for each of the three research questions. After the theoretical risk identification process, the aspects of practical risk awareness and risk handling will be analyzed. Based on the obtained insights and results, the author provides general recommendations in the form of a risk control matrix and establishes a holistic Cryptocurrency IC Framework. Finally, the conclusion in Chapter 5 summarizes and discusses this study and identifies its limitations.

## 1.3. Empirical Approach

The empirical approach employed to investigate the research questions was an online survey sent to a pre-determined group of companies accepting Bitcoin payments. Because of the small number of companies currently accepting Bitcoin payments in Switzerland, the selection process of the sample was not limited by any particular industry or geographical region within Europe.

The survey contains qualitative and quantitative questions about the Bitcoin transaction process, Bitcoin administration and the awareness as well as the evaluation of risk factors arising from the use of cryptocurrencies. Further, companies were also asked about the internal controls implemented, if any, for their operational dealing of cryptocurrencies.

The evaluation of the survey responses will be conducted in a descriptive manner. In this context, qualitative questions will be assessed using the concept of inductive content analysis, whereas quantitative questions will be evaluated using nominal, ordinal and interval scales.

## 2. Literature Review

### 2.1. Cryptocurrencies

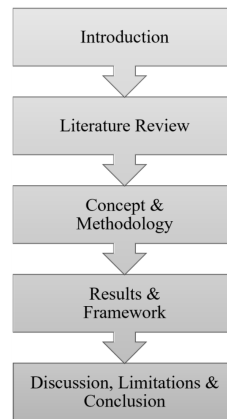
"There are 3 eras of currency: commodity based, politically based and now, math based."  
Chris Dixon, 2015

#### 2.1.1. Taxonomy

The story of forms of money can be traced back to the early ages of mankind when people traded goods for valuable commodities, such as seashells or cattle. Over time, several other kinds of payment instruments emerged like metals or coins, all of which were linked to their intrinsic value, depending on their weight and purity. (Dombrowski, 2014, p. 9f.)

The introduction of paper money simplified the handling and exchange of money, making transactions more convenient. The value of the paper could be derived in various ways, e.g., by pegging it to a scarce commodity like gold (the former "gold standard") or what is commonly known as "fiat money." The value of fiat money, such as the US Dollar, the British Pound, or the European Euro is determined by "the demand and supply of the nation's currency and its acceptance is enforced by the government through such means as declaring it legal tender" (Dombrowski, 2014, p. 12). This implies that its value depends on factors such as social convention or future expectations of economic developments (Franco, 2014, pp. 4-5).

According to Nian and Chuen (2015, pp. 7-8), "there are various socioeconomic forces that drive the demand for alternative currencies," such as political instability, technological progress, or inefficiencies. An accumulation of these factors might lead to dissatisfaction and consequently to a loss of trust in national currencies and the financial system



**Figure 1:** Structure of this Thesis; Source: Own illustration

**Table 1:** Comparison of Various Types of Currency; Source: Own illustration according to [Greene and Shy \(2014, p. 275\)](#)

|                       | Physical      | Digital          |
|-----------------------|---------------|------------------|
| Government backed     | Cash          | E-Cash           |
| Not government backed | Private Money | Virtual Currency |

as a whole. A prime example of this was the financial crisis of 2008. In order to establish an independent system that aims at tackling the above-mentioned causes for the demand of alternative currencies and the related discontent, Satoshi Nakamoto (2008, p. 1) proposed the concept of the first decentralized digital currency in a whitepaper entitled, "Bitcoin: A Peer-to-Peer Electronic Cash System."

Table 1 presents a classification of digital currencies as proposed by [Greene and Shy \(2014, p. 275\)](#).

As can be seen in Table 1, there is a general distinction between physical and digital currency. Additionally, there is a differentiation between currency that is or is not government backed. According to the recent definition of the US Government Accountability Office, "a virtual currency is generally considered as a digital unit of exchange that is not backed by governmental-issued legal tender" ([United States Treasury Inspector General for Tax Administration \(TIGTA\), 2016, p. 1](#)), whereby its "value [is] stored electronically in a device such as a chip card or a hard device in a personal computer" ([Bank for International Settlement \[BIS\], 2015, p. 4](#)). This declaration contradicts the basic property of e-cash, because it represents government-issued money; however, it is even stored in the form of bits on a chip, like a chip card or a mobile device ([Green & Shy, 2014, p. 275](#)).

To better understand a digital cryptocurrency and its structure, it is important to keep in mind that a digital currency (or "virtual currency" as it is also known) describes the "umbrella term," whereas "cryptocurrency" solely represents a subset with specialized properties ([Nian and Chuen, 2015, p. 6](#)). For instance, airline miles or tokens for online games can be classified as digital currencies as well ([Nian & Lee Kuo Chuen, p. 8](#)). Figure 2, reproduced from the [IMF \(2016, p. 8\)](#), presents an accurate taxonomy of virtual currencies.

Figure 2 can be summarized as follows: A cryptocurrency

- has no legal tender
- is a medium of exchange for "real-world" items or "real-world" money
- has no central authority that controls and regulates the money circulation
- uses mathematics, particularly cryptography, to validate transactions.

Using the example of Bitcoin and its blockchain, the following chapters will present an in-depth explanation of the structure and application of Bitcoin as a payment method.

### 2.1.2. Bitcoin

We begin this section by pointing out the general distinction made between the two forms of this term, which are commonly used in English- and French-speaking areas: "Bitcoin" refers to the network and its technology, while "bitcoin" refers to the unit of currency ([Gisler, 2015, pp. 9-10](#)).

#### Concept

The fundamental idea of Bitcoin is derived from the white paper entitled "Bitcoin: A Peer-to-Peer Electronic Cash System," written by the pseudonymous Satoshi Nakamoto in 2008. This paper illustrated the idea of a purely peer-to-peer (P2P) version of electronic cash without the involvement of any financial third parties ([Nakamoto, 2008, p. 1](#)). Bitcoin became the first decentralized digital currency, which implies that there is neither a person nor an institution who is backing, controlling, or regulating the currency ([Franco, 2014, pp. 3-4](#)). Instead, Bitcoin "is shared by all network nodes, updated by miners, monitored by everyone, and owned and controlled by no one" ([Swan, 2015, p. 1](#)).



Figure 1. Taxonomy of Virtual Currencies

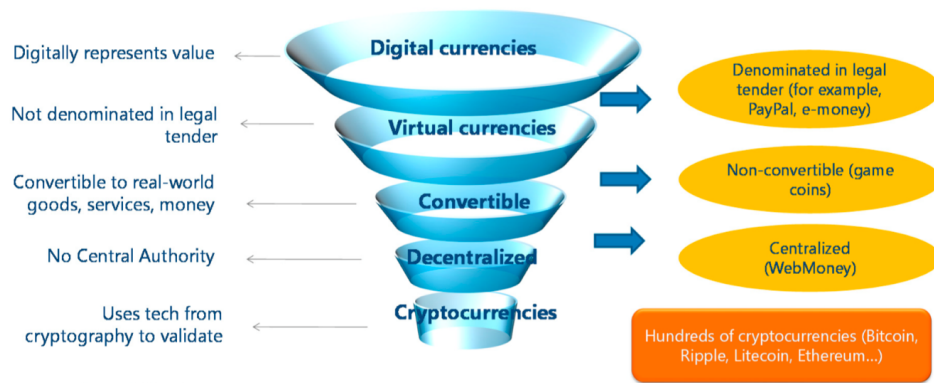


Figure 2: Taxonomy of Virtual Currencies; Source: IMF (2016, p. 8)

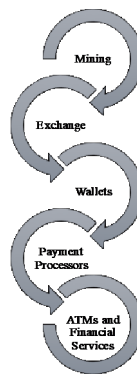


Figure 3: The Bitcoin Value Chain; Source: Own illustration according to Bitcoin Foundation (2017)

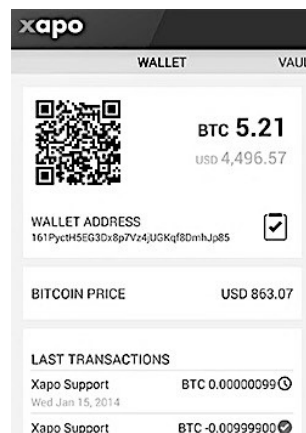


Figure 4: Bitcoin Wallet Screenshot; Source: <https://bitcoin.org/img/screenshots/xapo>

Nakamoto (2008, p. 1) presented the idea of "an electronic payment system based on cryptographic proof instead of trust." It aims at tackling the problems associated with a centralized financial system and further protecting buyers and sellers from fraud by means of the irreversibility of transactions as well as a P2P distributed timestamp server structure. Moreover, it creates the possibility of reaching the unbanked population and aims at eliminating the disproportional remittance and exchange fees involved in,

especially, small-amount transactions. For instance, when sending money to any African countries or when supporting any micro-financing projects, the remittance bears the risk of the transaction cost being higher than the amount originally transferred (Nian and Chuen, 2015, pp. 13-23). Additionally, the proposed approach tackles the so far unsolved double-spending problem in a distributed financial database without the necessity of an intermediary who charges (high) transaction fees for the validation, such as banks, credit card

companies, or other established payment transmitters like PayPal (Deloitte, 2016, p. 5; Franco, 2014, p. 6; Papadopoulos, 2015, p. 158).

According to Papadopoulos (2015, p. 155), bitcoins are the digital equivalent of cash, which is designed to transfer "economic value across the internet." Basically, the concept of Bitcoin can be compared with mining gold, whereby the globally available amount of gold would be equal to the limited availability of bitcoins (Tindell, 2013, 5. March). To further explain this comparison, the bitcoin "creation process" – the so-called "mining" – is conducted by a decentralized P2P network of computers working around the clock to solve difficult mathematical problems based on cryptographic hash algorithms instead of gold miners physically digging for gold (Antonopoulos, 2015, p. 175; Cusumano, 2014, p. 22).

When a mathematical problem is solved, the successful miner(s) are rewarded in the form of bitcoins for providing their processing power and the necessary energy. Finally, the earned as well as the previously spent bitcoins are stored "in a notional ledger held across many computers around the world" (the blockchain) and not in underground vaults as in the case of gold bars (Tindell, 2013, 5. March).

#### *The Bitcoin Value Chain*

As described in the previous section, the bitcoin creation process differs from the issuance of fiat money, whose supply is controlled and regulated by the central bank. In comparison, in the Bitcoin ecosystem, the blockchain, which is based on a P2P network of computers executing a determined algorithm (the Bitcoin protocol), takes charge of this responsibility. This implies two core properties of Bitcoin: first, there is no centralized regulatory instance that is responsible for these tasks, and, second, the supply of money is limited to 21 million bitcoins because of one (of many) requirements determined by the Bitcoin protocol (Bhaskar and Chuen, 2015, p. 53). Additionally, the mining process is defined through a predictable progressive growth rate, which will converge toward zero until the year 2140. Hence, the issuance as well as the mining rewards will be halved approx. every four years which is equivalent to every 210,000 blocks mined (Bhaskar and Chuen, 2015, p. 53).

In 2009, the first bitcoin was mined and as of January 20, 2018, approx. 16.8 million bitcoins have been issued, which is roughly 80% of all the potentially available bitcoins (CoinDesk, 2018, 20. January). When the newly emerged bitcoins are mined, they are first added to the blockchain and, subsequently, attributed to the successful miner(s). A miner has two options: on the one hand, he has the possibility to sell the bitcoins against fiat currencies or other cryptocurrencies on a specific cryptocurrency exchange platform like bitfinex.com or bittrex.com. On the other hand, the rewarded bitcoins can be transferred directly to his wallet and can thus be used for personal purposes, e.g., as a medium of exchange for real or digital goods and services or as an investment (de Jong, 2015, pp. 416-417). An overview of the Bitcoin value chain can be found in Figure 3.

In order to use and store bitcoins, regardless of how the

bitcoins are received – as a mining reward or bought on an exchange platform – one has to download a so-called "wallet." A wallet is an application, typically for smartphones, tablets, or desktop computers, through which an owner of bitcoins is able to store, send and receive Bitcoin funds (Bitcoin Foundation, 2017). Moreover, a wallet runs a key generation software that creates a public address (known as a wallet address) and stores the corresponding private key (Franco, 2014, p. 56). In this context, wallets are often described as "containers for private keys," because the bitcoins themselves are stored on the blockchain and not in the user's wallet, i.e., a wallet keeps only the corresponding private key to sign a transaction and transfer the property rights of a bitcoin (Antonopoulos, 2015, p. 86). More information about the transaction process is given in Section 2.1.3.2. A wallet application also enables the conversion of public addresses to QR codes, the management of incoming and outgoing payments, and displays the currently available funds (Sixt, 2017, p. 37). A screenshot of a wallet can be seen in Figure 4.

As described, Bitcoin payments can be sent or received by means of a wallet application. Therefore, the sender has to type in the public address (or scan the QR code) of the recipient and specify the number of bitcoins that should be debited from the account.

In order to simplify the implementation of Bitcoin acceptance, to overcome the general complexity of the technological novice, and the "bottlenecks in exchanging" into fiat currencies, merchants often use a payment processor (Papadopoulos, 2015, p. 161). This is a company that serves as a third-party provider and offers services for retailers to handle their payments by providing a technical infrastructure such as payment terminals, mobile apps, or e-commerce plugins. These tools are commonly able to quote prices in the national currency, directly convert bitcoins into the national currency, and instantly confirm the transactions and thus ensure that there is no risk of a payment default. All of this is done in exchange for a transaction fee amounting to 0.5-1% of the invoiced amount. This is cheaper than using credit cards whose fees are 2-3% per transaction. (Cusumano, 2014, p. 23; Franco, 2014, p. 43)

At the end of the Bitcoin value chain, there are the several applications of Bitcoin, which are mainly based in the payment and investment industry, for instance, making purchases, holding it as an investment asset, or making further investments in other currencies or decentralized autonomous organizations (DAOs). Due to the increasing acceptance but continued lack of know-how about the complicated procedures behind the entire value chain, Bitcoin-interested parties are also able to instantly exchange their (fiat) money in bitcoins more conveniently by way of a Bitcoin teller machine (BTM). They can then later transfer them to their wallet and are thus finally able to participate in the Bitcoin application universe (Cohen, 2014, pp. 18-19).

#### *Merchant Acceptance & Market Capitalization*

At the time of writing this thesis, there are more than 1500 established cryptocurrencies (cf. Worldcoinindex, 2018, 9.

February). According to Elfriede Sixt (2017, pp. 111f.), the alternatives to Bitcoin are called "altcoins," which differ in several ways, but primarily in terms of the underlying technological structure (blockchain and protocol), which influences the validation process, the money supply, and the purpose of the digital token issued (currency, smart contract, investment token, usage token, utility token, etc.).

In the recent years, an increasing number of merchants have started to accept Bitcoin payments. The payment processor Coinbase's advertisements state that "over 48,000 merchants" use its technical infrastructure to accept Bitcoin payments. This includes multinational companies from several industries, such as the online retailer overstock.com, the IT company Dell, and the online travel company Expedia (Coinbase, 2018, 12. February). Moreover, there are at least ten payment processors, apart from companies that do not use any payment processor. Unfortunately, there is no global unambiguous number of merchants who actually accept Bitcoin payments. For instance, the platform coinmap.org (optional registration platform for Bitcoin merchants) lists approx. 11,400 merchants who accept Bitcoin payments as on January 5, 2018 (Coinmap, 2018, 5. January).

In fact, as of January 20, 2018, Bitcoin had a total market cap of USD 214.2 bn., which is equivalent to 34.8% of the entire crypto market cap. Further, about 315,000 Bitcoin transactions had been conducted in the previous 24 hours (Blockchain, 2018, 20. January; Coinmarketcap, 2018, 4. February) – six times more than four years ago. Moreover, one bitcoin has a value of around USD 12,600 (Coinmarketcap, 2018, 4. February), which has increased by approx. 1400% during the last year.

### 2.1.1.3. Blockchain

In recent times, blockchain technology has attracted more and more industries. The application of the technology goes far beyond the scope of cryptocurrencies. Massimo Di Pierro (2017, p. 92), professor at the School of Computing of DePaul University, states that the problems of authentication, integrity, and non-repudiation are solved by this disruptive technology because it deals with the "problem of creating a distributed storage of timestamped documents where no party can tamper with the content of the data or the timestamps without detection."

The decentralized database system enables the emergence of new use cases due to its ability to assign things to identities in a tamper-proof and decentralized manner, such as a decentralized notary, e.g., the attribution of physical goods like real estate, gold, or diamonds to digital identities (Deloitte, 2016, p. 5). Particularly, in the world of intellectual property, the double-spending issue represents a huge problem because of the "copy and paste" possibilities. Based on the fact that there is no database documenting the purchase/sale of a digital token (such as a song, for example), the token can be shared in an unlimited manner (Vigna and Casey, 2016, p. 120).

This vulnerability of double-spending can be eliminated by means of the decentralized consensus process of the Bit-

coin blockchain (Underwood, 2016, p. 15). In general, "a double-spend attempt occurs when a user tries to spend some funds twice," which plays especially a significant role in the financial system because all financial institutions must reject these attempts (Franco, 2014, p. 6). In a centralized system, this "is relatively straightforward [...] because transactions are recorded in a central database and future spending attempts are checked against this database first" (Franco, 2014, p. 6). Controversially, in a "decentralized system, many copies of the database are shared among the peers, and keeping a consistent state of the database is a difficult computational problem" (Franco, 2014, p. 6).

According to Antonopoulos (2015, p. 3), the Bitcoin network exhibits, by means of its underlying technology (the blockchain), four key innovations functioning in a "unique and powerful combination":

- A decentralized mathematical and deterministic currency issuance (Mining)
- A decentralized P2P network
- A decentralized transaction verification system
- A public transaction ledger

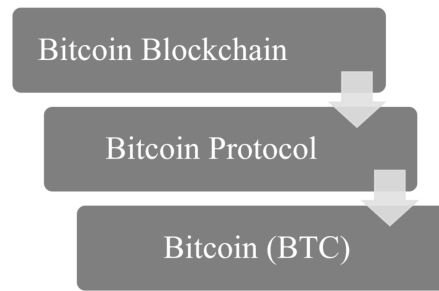
In order to explain these innovations, this chapter is divided into three subsections: Structure, Transactions Process, and the Decentralized Consensus of a Transaction (Mining).

#### Structure

The blockchain is the underlying technology that operates a cryptocurrency and thus represents the fundamental prerequisite for its existence. Cryptocurrency experts like Vigna and Casey (2016, p. 120) describe the Bitcoin blockchain as "bitcoin's central nervous system." Some confusion might arise from the fact that the term Bitcoin is used for all three layers in the technology stack of the Bitcoin blockchain, i.e., "Bitcoin" refers to the blockchain, its protocol, as well as the available currency and tokens issued (Franco, 2014, pp. 18-19). However, Figure 5 presents a general link between these three layers.

The Bitcoin blockchain is essentially a database that is represented by its nodes and "replicated as many times as there are nodes" (Morettini, n.D.). These nodes provide their central processing unit (CPU) to the Bitcoin network to run a software known as the Bitcoin protocol. This is why Morettini (n.D.) states that the Bitcoin blockchain can even be described as a "supercomputer formed by the combination of the CPUs [...] of all its nodes." In this context, it is important to understand that the Bitcoin blockchain is not a tool for the transmission of bitcoins; it is rather "an authoritative public record that records the chain of title for any current bitcoin holdings" (Lloyds, 2015, p. 7). This implies that from a technical point of view, a blockchain is a database allowing data collection in the form of data packages which can be easily accessed, updated, and managed (Rouse, 2017).

From an application-oriented, comparable, and conceivable view, the Bitcoin blockchain is often considered as a



**Figure 5:** Layers in the Bitcoin Blockchain Technology Stack; Source: Own illustration according to Swan (2015, p. 1)

"general ledger," because each transaction that has ever been conducted since January 3, 2009, is recorded in this publicly accessible and non-reversible transaction ledger (Sixt, 2017, p. 30; Franco, 2014, p. 95). In general, a blockchain can be compared with an endless table containing three columns: the transaction's timestamp, the transaction's details, and the corresponding hash (Di Pierro, 2017, p. 93).

The Bitcoin protocol, which is executed by the nodes, takes charge of the proof and verification process of transactions determined in a decentralized consensus process (Vigna and Casey, 2016, p. 124). The underlying software algorithm (the Bitcoin protocol) determines the steps that any node has to execute in order to accomplish a general consensus for a transaction, i.e., either to validate and confirm or to deny a transaction (Sixt, 2017, pp. 1-2). Additionally, the protocol defines the properties of cryptocurrency, such as the (limited) amount of coins issued or the explicit details (halving time, coins per block, etc.) about the (deterministic) coin issuance. In general, it can be stated that the Bitcoin protocol sets the rules and the framework conditions for a cryptocurrency.

Thus the existence of a cryptocurrency token (such as one bitcoin) can be attributed to a database (the blockchain), which operates the appropriate software (the protocol). Each cryptocurrency consists of at least these three layers (Swan, 2015, p. 1).

#### Transaction Process

The transaction process can be generally divided into two steps: the authorization process by means of digital signatures and a decentralized verification and validation process (mining).

A bitcoin is a unique number that results as the solution of a cryptographic puzzle (Di Pierro, 2017, p. 93). Thus, a transaction is the change of ownership of this number, which in turn requires the authorization of possessing and transferring this number, i.e., a single bitcoin or its fractions (Antonopoulos, 2015, p. 18). However, the question is how a transaction works and, in turn, how the ownership of a bitcoin changes. Bitcoin's inventor Satoshi Nakamoto (2008, p. 2) defined an electronic coin as "a chain of digital signatures." Hence, a transfer of the ownership and control of a coin is achieved "by digitally signing a hash of the previous transaction and the public key of the next owner and adding

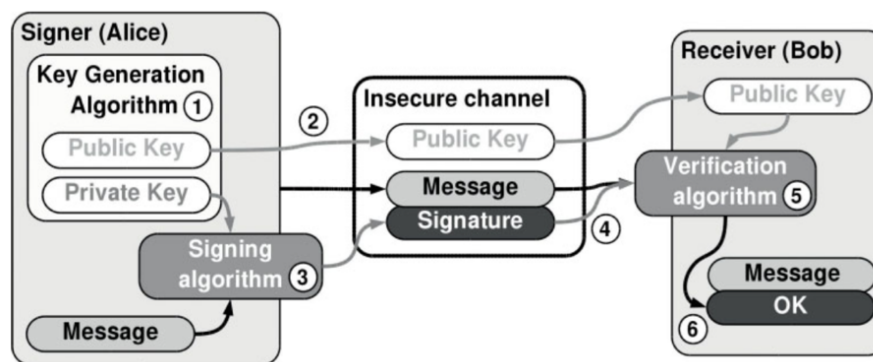
these to the end of the coin" (Nguyen, 2017, p. 1). In this context, the blockchain stores this chain of digital signatures (Lloyds, 2015, p. 7).

This sounds extremely complicated at first glance, but it can be simplified in the following way: In the Bitcoin network, the transfer of bitcoins is authorized and initiated by digitally signing a transaction. This digital signature process is based on a public-private key pair, which is based on its cryptographic origin, or, to be more specific, to the calculation of elliptic curves that are based on the Elliptic Curve Digital Signature Algorithm (ECDSA) (Sixt, 2017, p. 37). Nevertheless, although the public key can be compared with a bank account number, the private key serves as the signature unlocking this bank account (Franco, 2014, p. 56). The possession and control of a bitcoin (or its fractions) is the same as having knowledge about the private key that is mathematically linked to the public address/key (Lloyds, 2015, p. 7).

For instance, let's say a signer called Alice would like to buy a coffee at Bob's café (the recipient). To begin, Alice needs a public-private key pair that is usually automatically generated by her Bitcoin wallet. She then sends her public key to the recipient (2). In the meantime (3), she uses her private key to digitally sign the transaction terms (called a "message" in Figure 6) containing the transaction details (i.e., price and terms) along with a timestamp (Di Pierro, 2017, p. 93; Goldman Sachs, 2016, pp. 8f.). When the transaction is transmitted (4) to Bob, he is able to verify it using Alice's public key (5). If the verification goes through, he knows that Alice has triggered the payment and that she has transferred her ownership rights of the bitcoins to him (6). An overview of this digital signature process can be seen in Figure 6 above. In practice, these "technical" steps are not conducted by Alice or Bob but by their wallet applications, which are linked to the Bitcoin protocol. Finally, the signatures are proofed and verified by the nodes of the Bitcoin blockchain (Franco, 2014, pp. 56-57).

To sum up, a digital signature serves two purposes: "(1) to verify the transaction to be sent and validated to the Bitcoin network, and also (2) to confirm her assent to the transaction" (Nguyen, 2017, p. 2). When a transaction is signed, the verification and validation process begins by generating a computational hash, which is conducted by a decentralized network of miners in a process known as the "mining process", cf. Figure 7.





**Figure 6:** Digital Signature Process; Source: Franco (2014, p. 57)

### *The Decentralized Consensus of Transaction (Mining)*

This mining process is conducted by all network nodes or "miners," who are essentially electronic bookkeepers. In general, Bitcoin transactions are similar to the entries in a double-entry bookkeeping ledger (see Figure 8). On one side, there are the credits on a Bitcoin account (the inputs) resulting from one or more incoming Bitcoin payments or mining earnings. On the other side of the ledger are the debits (the outputs) that equal to the Bitcoin payments made. In this context, the inputs need not to be equal to the outputs, with the difference being the transaction fees for the miners. (Antonopoulos, 2015, pp. 18-19)

When a transaction is signed, it is first shared among the network nodes (Antonopoulos, 2015, p. 25). In order to review and validate the transaction, the miners include all transactions (since the last mined block) to the next block of the blockchain, which has to be subsequently mined. Now the mining process begins: Mining is based on a cryptographic concept where nonces are randomly guessed, until the right one is found – the so-called "proof-of-work" (Sext, 2017, p. 40). The nonce is a random number used for the verification of a unique hash calculated using cryptographic means (Nofer et al., 2017, pp. 183-184).

From a more comprehensive view, Bitcoin mining can be described as a competitive Sudoku "that resets every time someone finds a solution and whose difficulty automatically adjusts so that it takes approx. ten minutes to find a solution" (Antonopoulos, 2015, p. 26). The successful miner who solves the Sudoku (or the algorithm) and finds the proof-of-work, shares this solution with the other network nodes. The other miners verify the result and, subsequently, propagate the new block in the network (Antonopoulos, 2015, p. 26).

If this new block is validated, it will be "sealed" and recorded in the blockchain while the miners go on to solve the algorithm of the next block (Antonopoulos, 2015, p. 27). Furthermore, the timestamp and the nonce are added to the mined block and thus a blockchain can be compared with a table containing three columns: the transaction's timestamp, the transaction's details, and the corresponding hash (Di Pierro, 2017, p. 93). In order to make this database tamper-proof, the hash of the previous block is included as

well. Therefore, it is not possible to change any transactions that have already been conducted, verified and confirmed without changing the entire blockchain, because any change would immediately affect the hash values calculated (Nofer et al., 2017, pp. 183-184). This implies that no one can counterfeit the blockchain without the network noticing it, because each peer keeps a copy of the entire common asset ledger (Franco, 2014, p. 8; Morettini, n.D.). Figure 9 provides an overview of an example of a blockchain.

Referring again to our previous example, Alice gets her first payment confirmation after the block containing her transaction is sealed. If the next block is mined, Alice will get her second confirmation and so on. This implies that each block mined on top of the block containing the initial transaction represents an implicit confirmation and thus it will be exponentially harder to reverse or manipulate a transaction. Usually, a block with more than six confirmations is considered as "irrevocable" because it would cost an extremely large amount of computational power to recalculate six blocks (Antonopoulos, 2015, p. 28).

As already described, the miners provide their computers' power (CPUs) to the Bitcoin network in order to execute the Bitcoin protocol, which determines the steps a computer has to execute to validate or deny a transaction to the blockchain (Böhme et al., 2015, p. 215; de Jong, 2015, p. 417). In exchange for providing their processing power, the miners get paid for their "work" through incentives or "mining rewards," which is in the form of newly issued bitcoins (Gisler, 2015, pp. 13-14). Every ten minutes, new bitcoins are issued due to the period of the validation process (Vigna and Casey, 2016, p. 132). See Figure 10 that outlines this entire process.

The new issuance of bitcoins therefore occurs six times per hour and 144 times per day, or, in other words, 144 blocks are mined per day. Due to the fact that the mining rewards are halved every four years, cryptocurrency experts like Antonopoulos (2015, p. 3) speak about a "deterministic currency issuance" in case of Bitcoin. Currently, there are approx. 500,000 mined blocks and the mining reward per block is about 12.5 bitcoins (Blockchain, 2018, 20. January).

In summary, it can be stated that the mining process fulfills two interrelated core tasks: on the one hand, the is-



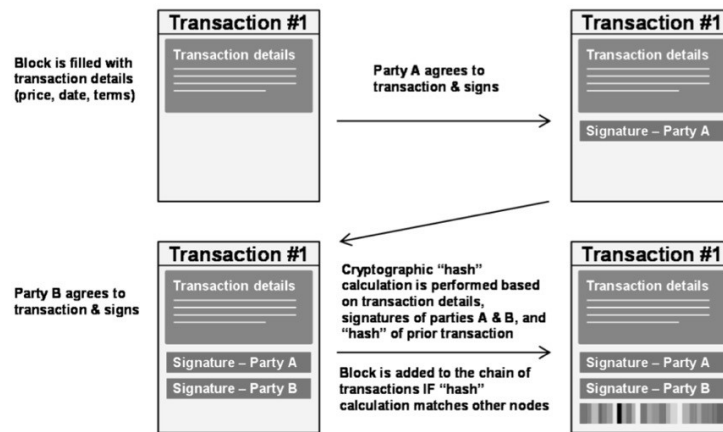


Figure 7: Transaction Process; Source: Goldman Sachs (2016, p. 8)

| Transaction as Double-Entry Bookkeeping                                                                                                                                                 |          |                |          |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|----------------|----------|
| Inputs                                                                                                                                                                                  | Value    | Outputs        | Value    |
| Input 1                                                                                                                                                                                 | 0.10 BTC | Output 1       | 0.10 BTC |
| Input 2                                                                                                                                                                                 | 0.20 BTC | Output 2       | 0.20 BTC |
| Input 3                                                                                                                                                                                 | 0.10 BTC | Output 3       | 0.20 BTC |
| Input 4                                                                                                                                                                                 | 0.15 BTC |                |          |
| Total Inputs:                                                                                                                                                                           | 0.55 BTC | Total Outputs: | 0.50 BTC |
| <div> <div>Inputs</div> <div>0.55 BTC</div> </div> <div> <div>Outputs</div> <div>0.50 BTC</div> </div> <div> <div>Difference</div> <div>0.05 BTC (implied transaction fee)</div> </div> |          |                |          |

Figure 8: Transaction as Doubly-Entry Bookkeeping; Source: Antonopoulos (2015, p. 19)

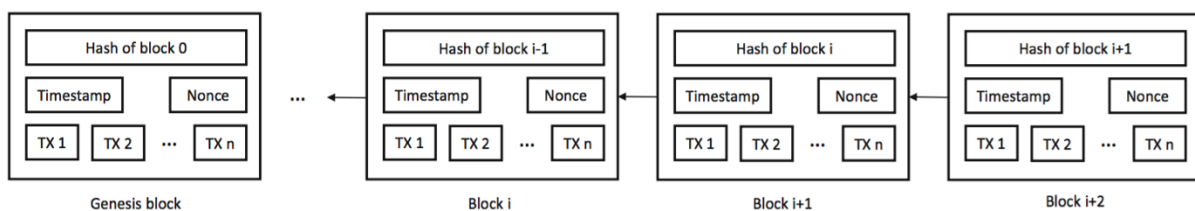


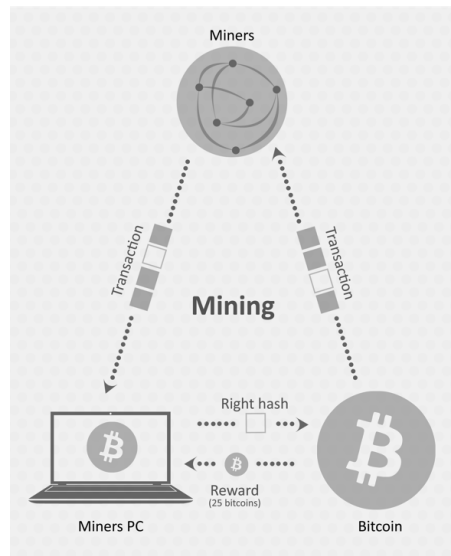
Figure 9: Example of a Blockchain; Source: Nofer et al. (2017, p. 184)

suance of new bitcoins, which is quite similar to a central bank printing new money, and, on the other hand, the validation of transactions, which is its main purpose (Antonopoulos, 2015, pp. 25; 175-176). An overview of the entire mining process can be found in Appendix 1 – The Mining Process.

#### 2.1.4. Regulation

Notwithstanding, Bitcoin has gained popularity no just as a form of currency in the context of e-commerce. Particularly in consideration of regulatory issues, cryptocurrencies

have in the meantime caught the attention of countries and communities of states as well. The increasing market capitalization and discussions about the classification of virtual currencies have fostered the consideration of virtual currencies as a subject of regulation; for instance, in April 2017, Japan announced that it officially recognizes Bitcoin as a legal tender (Gautham, 2017). In comparison, there are also countries such as Bolivia, Vietnam, and Bangladesh that have completely banned Bitcoin and declared cryptocurrencies as illegal (Swan, 2015, p. 7).



**Figure 10:** The Bitcoin Mining; Source: <http://en.bitcoinwiki.org/Mining>

In 2012, the European Central Bank (ECB, 2012, p. 13) classified Bitcoin as a virtual currency with the following definition: "A virtual currency is a type of unregulated, digital money, which is issued and usually controlled by its developers, and used and accepted among the members of a specific virtual community." Until now, Bitcoin is still not specifically regulated in the European Union (EU) including Switzerland (Piller, 2017, p. 9); however, in recent weeks, there have been many proposals to consider virtual currencies at least in the national regulations around Anti-Money-Laundering (AML) and Know-Your-Customer (KYC) issues (Kollewe, 2017, 4. December). This is perhaps mainly due to the significant price development (cf. Appendix 2) and the remarkable market capitalization during the last year. Therefore, a greater number of voices are now demanding a stronger regulation of Bitcoin, especially in consideration of money laundering issues (Berschens et al., 2017, 19. December; Kollewe, 2017, 4. December).

## 2.2. Enterprise Risk Management & Internal Control

### 2.2.1. Definitions & Relevance

Risks are dynamic, fluid, and extremely interdependent (Lam, 2014, p. 51). In today's highly volatile and dynamic environment, a company requires an integrative approach to manage its risk portfolio instead of single detective controls that are implemented and monitored by separate business units (Lam, 2014, p. 51). For instance, the implementation of a new payment method, such as the acceptance of cryptocurrencies, does not solely influence the accounting and controlling department as an additional payment option that has to be considered in the revenue recognition and consolidation. There might even be some fundamental adaptations due to emerging risks in the underlying IT infrastructure/security, operational dealings, as well as the compliance of new regulations by accepting these payments. Therefore,

an integrative approach across several business units is required.

The ERM function takes charge of this responsibility by establishing "firm-wide policies and standards, [coordinating] risk management activities across business units and functions, and [providing] overall risk monitoring for senior management and the board" (Lam, 2014, p. 51). The objective of ERM is to ensure that the company's management is able to effectively deal with uncertainty and its associated risks. In order to achieve this, the approach aims at ensuring the effective running of business operations, accurate reporting and compliance with regulations to prevent any events that might damage the company's reputation (COSO, 2004, p. 1).

The concept of internal control (IC) is an integral part of an entity's risk management. In general, IC is defined as "a process, effected by an entity's board of directors (BoD), management, and other personnel, designed to provide reasonable assurance regarding the achievement of objectives relating to operations, reporting, and compliance" (COSO, 2013, p. 3). This implies that IC does not only describe the detective and the corrective controls associated, it also proposes an integrative approach including directive and preventive controls to ensure that no misconduct or adverse event can occur that might negatively affect the company's objectives (Ruud and Jenal, 2005, p. 456).

For this purpose, the implementation of the COSO IC Framework provides a proper method of solution, established by the Committee of Sponsoring Organizations of the Treadway Commission (COSO). This risk management framework outlines components, principles, and factors that are necessary to effectively manage risks (COSO, 2015, p. 1). However, "although risk management frameworks can effectively identify the types of risks that modern businesses must control, these frameworks are largely silent about how specific duties should be assigned and coordinated within the orga-

nization" (of *Internal Auditors*, IIA, p. 1). Therefore, the proper assignment of duties and responsibilities related to IC within an organization's governance is a prerequisite. In this context, the Three Lines of Defense (TLoD) model provides a simple and systematic approach to improve the effectiveness of a company's risk management (of *Internal Auditors*, IIA, p. 2).

### 2.2.2. Internal Control

#### *In an Organization's Governance*

By describing IC's responsibilities and further differentiating it from the internal audit function within a corporation, the TLoD model, established by the European Confederation of Institutes of Internal Auditing (IIA) and the Federation of European Risk Management Associations, is being increasingly used and generally accepted as an institutional framework (Casari, 2017, p. 17; Hunziker, 2015, p. 52). This framework is employed to address "how specific duties related to risk and control could be assigned and coordinated within an organization" with the objective of increasing the likelihood of achieving the company's objectives (COSO, 2015, p. 1). It can be seen as a set of guidelines for how risk and control duties must be allocated and executed within an organizational structure. Moreover, such effective coordination of responsibilities prevents a duplication of efforts and controls, which in turn fosters a more effective risk management approach (of *Internal Auditors*, IIA, p. 1).

The TLoD model is based on the assumption that "three lines" within an organization are necessary for the effective management of risk and control (cf. Figure 11):

- 1st Line of Defense: Operating management that is responsible for owning and managing risk and control
- 2nd Line of Defense: Management support that takes care of monitoring the risk and control
- 3rd Line of Defense: Independent assurance concerning the management of risk and control

As can be seen in Figure 11, the senior management and the BoD as well as external parties such as the external auditors or regulators play integral roles (Hunziker, 2015, pp. 52-53). Whereas the senior management is responsible for selecting, developing, and evaluating the IC system, the BoD will manage and overview the entire process (COSO, 2015, p. 4). Both groups are primary stakeholders, served by the three lines. It is their "responsibility and accountability for setting the organization's objectives, defining strategies to achieve those objectives, and establishing governance structures and processes to best manage the risks in accomplishing those objectives" (of *Internal Auditors*, IIA, p. 3). Additionally, they are in the best position to help ensure and actively support the implementation of the TLoD (of *Internal Auditors*, IIA, pp. 2-3).

External auditors, regulators, and other external parties are outside the organization's structure. Even still, they are able to influence the organization's government and control

structure, e.g., by setting requirements to strengthen the control or by performing independent assessments. This implies that an external body can become an additional line of defense by "providing assurance to the organization's shareholders, including the governing body and senior management." (of *Internal Auditors*, IIA, p. 6)

To sum up, the TLoD framework can be described as a prerequisite but also as a complimentary framework to the COSO IC Framework in order to assign and allocate duties within an organization and to thus ensure proper risk management.

#### *COSO Internal Control – Integrated Framework*

In 1992, the COSO issued a report presenting the first COSO IC Framework (Arwinge, 2012, p. 37). In general, the US-based organization is dedicated to improving a companies' financial reporting quality, IC and Corporate Governance (Arwinge, 2012, p. 37). Over time, the COSO has developed and further modified several frameworks that concretize ERM and IC (Graham, 2015, p. 2). These frameworks outline components, principles, and factors that are necessary to effectively manage risks. Moreover, they provide evaluation tools for an organization to assess its IC (COSO, 2015, p. 1; Graham, 2015, p. 2) and therefore represents a worthwhile tool for "directors, managers, auditors, regulators, investors and other concerned stakeholders" (Arwinge, 2012, p. 37).

In the literature on this subject, there is some confusion about the COSO IC and the COSO ERM Frameworks, because the COSO has also published a COSO ERM Framework in 2004. This framework explicitly addresses control mechanisms for the achievement of strategic goals, which thus additionally provide "a more robust and extensive focus on the broader subject of enterprise risk management" (COSO, 2004, p. 5; Hunziker, 2015, p. 36). This implies that the ERM framework intends to expand the initial COSO IC Framework and to establish a more comprehensive framework. However, Buber and Holzmüller (2009, p. 106) states that the ERM and IC Frameworks are almost identical. This statement can be further justified by the fact that the COSO ERM Framework already encompasses the COSO IC Framework. According to the COSO (2013, p. ii) itself, both frameworks are "intended to be complimentary" and they do not supersede each other. For the further analysis, the author focuses on the COSO 2013 IC Framework, shown in Figure 12.

The COSO IC Framework enables "organizations to effectively and efficiently develop and maintain systems of internal control that can enhance the likelihood of achieving the entity's objectives and adapt to changes in the business and operating environments" (COSO, 2013, p. i).

In order to achieve this, the COSO IC Framework comprises three divisions (see Figure 12): objectives (the three columns), components (the five rows), and organizational structure (the third division). The first and second division are directly related to each other because the entity's objectives – which can be divided into operational, reporting, and compliance categories – require all five components for their

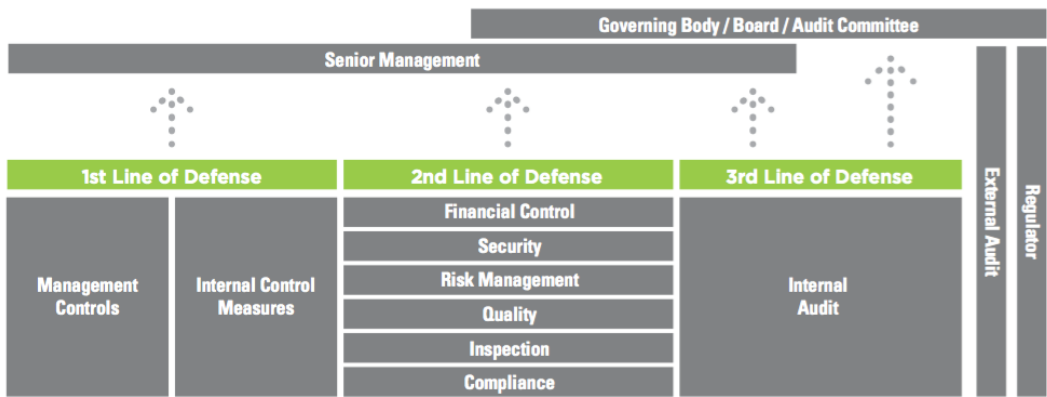


Figure 11: The Three Lines of Defense; Source: COSO (2015, p. 2)

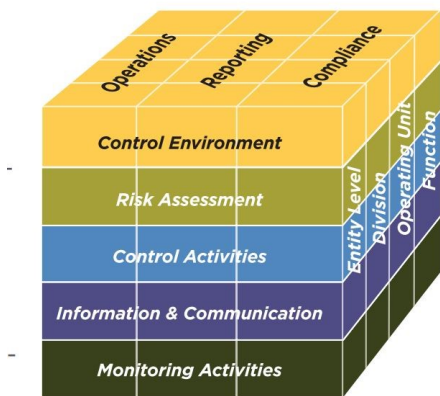


Figure 12: COSO 2013 Internal Control - Integrated Framework; Source: COSO (2013, p. 6)

success. Moreover, IC is relevant to the entire organization (third division), i.e., from the group level to each division, including any business unit and function. (Bungartz, 2017, p. 20; COSO, 2013, p. 6)

The five interrelated key components of IC are: control environment, risk assessment, control activities, information & communication and, monitoring activities (Arwinge, 2012, p. 37; Graham, 2015, p. 2). In a brief summary, they can be described in the following way:

- **Control Environment:** The control environment defines the basics of the IC within an organization. In this first stage, the "tone at the top" of a company is set when the responsibilities of the BoD and the senior management are negotiated and determined. This has a considerable impact on the working atmosphere within the company, the daily business operations, but also on the handling of risks (Ruud et al., 2008, p. 939). In consideration of factors such as integrity and ethical values and organizational structure and management philosophy, the control environment compromises the establishment of a "sound control environment." This sets the context standards for ethical behavior, corporate culture, and the basis for executing IC across an organization, including the expected

standards of conduct (Arwinge, 2012, p. 43).

- **Risk Assessment:** When the company's objectives and its corresponding control environment are defined, the risks that might negatively affect these can be identified by means of a risk assessment (Ruud et al., 2008, p. 940). Generally, emerging risks can be from internal or external sources. In this context, the risk assessment represents a dynamic and iterative process that evaluates the risks identified according to their likelihood and potential damage (COSO, 2013, p. 4). This process plays an increasingly important role in today's dynamic and fast paced environment and finally provides the basis for decisions concerning how and which risks have to be managed (Bungartz, 2017, p. 39).
- **Control Activities:** Based on the risk assessment performed, the management decides which control measures (preventive, directive, detective, or corrective) have to be implemented to properly address the risks identified (Hunziker, 2015, p. 32; Pfaff and Ruud, 2013, pp. 73-74). For this decision process, the development of a risk control matrix can be quite useful. According to Pfaff and Ruud (2013, p. 85), a risk control matrix can be employed as tool in order to deal

with, present, and document the risks identified. Table 2 presents a simplified version of a risk control matrix. Depending on the organization's size, the matrix could be more sophisticated to give a more comprehensive understanding of the risks. However, a risk control matrix also enables greater transparency for all involved parties, ranging from employees to external auditors, which improves risk awareness and the appropriate development and adaption of internal controls (Pfaff and Ruud, 2013, p. 86). These implemented control measures may include a range of manual activities like the review of necessary documents as well as automated controls such as software-based journal entry testing. In general, control activities are actions established through policies and procedures "that help ensure that management's directives to mitigate risks to the achievement of objectives are carried out" (COSO, 2013, p. 4). These control activities may be located and further performed throughout the entire organization, i.e., at all levels of the entity and in all business processes (Arwinge, 2012, p. 46).

- **Information & Communication:** Information obtained from internal or external sources serves as the basis for the communication of internal responsibilities (COSO, 2013, p. 5). In order to ensure that the control activities determined are implemented and effective, the necessary and important information needs to first be identified. Afterwards, this information should be communicated in a comprehensive way, so that it builds the basis for every employee's actions with regard to management and IC in the company (Ruud et al., 2008, p. 940). This implies that information and communication are highly dependent and, moreover, are critical to help achieve a company's objectives (Casari, 2017, p. 23).
- **Monitoring Activities:** Finally, a company has to assess and ensure the quality of the internal controls implemented over time (Graham, 2015, p. 173), especially given that a company's internal controls have to be continuously adjusted to external and internal changes (Ruud et al., 2008, p. 940). Therefore, the implementation process and the effectiveness of the internal controls require the appropriate monitoring activities to ensure their presence and functioning. These activities may be performed in the form of ongoing evaluations, separate evaluations, or a combination of both (COSO, 2013, p. 5).

In addition to the framework proposed, the COSO has stipulated 17 principles (cf. Appendix 3) aimed at formalizing the above-mentioned concepts "in a more comprehensive manner" (Casari, 2017, p. 23). These principles have been directly derived from the five key components, and, according to COSO (2013, p. 6), effective IC can be achieved through their implementation and application.

### *A Regulatory Perspective*

The COSO IC Framework represents a key conceptual and practical framework for several groups of stakeholders (Arwinge, 2012, p. 39). Specifically in relation to the emerging issues around Corporate Governance, the framework becomes increasingly important because it serves the fundamental principles (Hunziker, 2015, p. 34). Even though there is no obligation to design and implement IC within an organization, there is certainly a tendency toward stronger regulations (Ruud and Jenal, 2005, p. 455). According to Hunziker (2015, p. 39), this phenomenon is an implication of the negative developments of economies and the several corporate scandals in recent times, such as the Enron or the Dotcom scandal.

The US can be cited as a model example in the context of more stringent regulations. Following the implementation of the Sarbanes-Oxley Act (SOX) in 2002, an organization's management is responsible for IC and has to make a statement with regard to the appropriateness of internal control over financial reporting (ICOFR) (Ruud and Jenal, 2005, p. 455). Additionally, SOX Section 404 prescribes that the financial statement auditor "must assess and report on both management's assessment of the ICOFR and the operating effectiveness on the ICOFR" (Arwinge, 2012, p. 65). The SOX regulation applies to all companies, including foreign subsidiaries, which are listed by the US Securities and Exchange Commission (Casari, 2017, p. 27; Westerhausen, 2005, pp. 100-101). Therefore, IC represents an integral part of the US' mandatory law and, moreover, needs attestation by an auditor.

In order to improve the Corporate Governance in the EU, its commission also adopted similar guidelines in 2006. The new guidelines bear some similarities to the SOX and aim at harmonizing the regulation of IC across Europe. However, because there are no further European regulations, the respective national laws are applicable (Hunziker, 2015, p. 43). For instance, in Switzerland, the Swiss Code of Obligations (CO) further specifies the role of IC. According to art. 728a, paragraph 1 cipher 3 CO & art. 727, paragraph 1 CO, the external auditor has the duty of examining whether "there is an internal system of control," with all public companies as well as any economically significant company being subject bound by these regulations (Ruud et al., 2008, p. 938). Moreover, the CO states that the BoD has the "non-transferable and inalienable duties" of the company's management and organization, which inter alia includes the organization "of the accounting, financial control and financial planning systems as required for management of the company" (art. 716a, paragraph 1 ciphers 1-3 CO). While the CO only mentions the existence of IC, the Swiss Code of Best Practice for Corporate Governance outlines further recommendations with regard to IC (Ruud and Jenal, 2005, p. 455). However, there are still no legally binding requirements only recommendations concerning IC design and implementation in the EU.



**Table 2:** Simplified Version of a Risk Control Matrix; Source: Own illustration according to Pfaff and Ruud (2013, p. 85)

| Objectives | Risks | Control Activities |
|------------|-------|--------------------|
| ...        | ...   | ...                |

2.2.3. Failures and Limitations

Although the implementation of the COSO IC Framework enhances the likelihood of achieving the entity’s objectives, one has to keep in mind that IC gives only a reasonable and not absolute assurance (Graham, 2015, p. 15). For instance, a bad judgment made on the basis of a simple error or bias or an external event might negatively affect an entity’s operational goals, which cannot be prevented by IC. To sum up, even an effective IC can experience failure due to the following reasons:

- Human error: If the people who have implemented and executed IC make errors, which, in turn, lead to control failures
- Management overrides: A manager is able to override IC control for selfish purposes
- Collusion: The ability to circumvent IC by means of secret agreements (COSO, 2013, p. 9, Graham, 2015, pp. 15-16; Ruud and Jenal, 2005, p. 456)

Hence, an organization’s management should be aware of these vulnerabilities "when selecting, developing, and deploying controls that minimize, to the extent practical, these limitations" (COSO, 2013, p. 9).

3. Methodology: Approach

3.1. Research Concept

According to Brace (2013, p. 6) the definition of the research objective represents the first step of an empirical analysis. The ultimate aim of this study is the development of an appropriate framework and providing a set of recommendations to ensure the proper operational handling of cryptocurrencies. Therefore, the study identifies the following three research questions (cf. Chapter 1.1):

- What are the potential risks and resulting threats of accepting cryptocurrencies as a payment method?
- Are the companies and the responsible employees aware of these risks, and how do they evaluate and address them?
- Are there any general control activities and recommendations necessary to ensure an appropriate risk management in the handling of cryptocurrencies?

In order to answer these questions, a research concept was developed based on a combination of literature and empirical analysis. Based on the literature review on cryptocurrencies (cf. 2.1) and ERM (cf. 2.2), the author derived and identified the potential risks of Bitcoin payments and divided them

into operational, reporting, and compliance risks in Chapter 4.1. The risk identification approach follows the COSO IC Framework (cf. 2.2.2.2), wherein the subsequent empirical analysis focuses on the stage of the risk assessment and the corresponding control activities of the COSO IC Framework. The results obtained of the risk identification provide, on the one hand, the answer to the first research question and, on the other hand, the basis for the empirical analysis that follows.

In the second part of the research concept, the empirical analysis (cf. 4.2) is conducted starting with an analysis of the general payment process. This is important to understand if there is any third-party intermediary (i.e., a payment processor) involved that already covers any risks and, moreover, to comprehend how the company deals with the incoming funds. Afterwards, the explicit administration of Bitcoin payments in a company will be addressed, which focuses on private key storage and its corresponding responsibilities. Further, the survey participants were tasked with conducting an assessment of the potential risks that were previously identified with regard to Bitcoin payments. Their responses will help answer the research question of to what extent companies are aware of and how they evaluate risk potential and severity. Finally, the current risk measures – provided that there are internal controls implemented – will be analyzed.

In the third and last step, the paper outlines specific recommendations in the form of a risk control matrix, which is based on a summary of the previous results (cf. 4.3.1). Furthermore, the author presents a more general and holistic approach, called the "Cryptocurrency IC Framework" (cf. 4.3.2) that was established following the COSO IC Framework. The framework integrates the results of this thesis. By combining the Cryptocurrency IC Framework and the developed risk control matrix, this thesis presents the recommendations necessary in order to ensure the appropriate operational management of cryptocurrencies.

3.2. Research Method

A research method can be defined as a "system for collecting information" that should be based on an "architecture that serves a theme or multiple themes" in order to yield a well-constructed data (Azzara, 2010, p. 16; Sue and Ritter, 2007, pp. 1-2). For the empirical analysis of this thesis, the author decided to employ an online survey with integrated display logics and redirection functions (see "survey" in Appendix 5). The survey required the participants to answer a series of questions within a certain time frame, which is necessary to ensure a considerable response rate and enhance the representativeness of the sample (Azzara, 2010, p. 25).

There are, of course, many different ways to conduct a survey, e.g., face-to-face interview, phone interview, written

interview, or online interview (Blasius and Baur, 2014, p. 662). An online questionnaire is an appropriate research tool in case of a fairly large sample that is geographically distributed (Blasius and Baur, 2014, p. 662). Moreover, Sue and Ritter (2007, pp. 13-14) state that the online option is well suited to sensitive questions because of its anonymity, and it prevents any bias in the responses through the absence of an interviewer (Brace, 2013, pp. 23, 26).

The ultimate goal of a questionnaire "is to meet research objectives by obtaining valid data from respondents who are properly screened and qualified" (Azzara, 2010, pp. 18-19). Therefore, a representative sample for the entire population has to be identified (see Chapter 3.4). In addition to the selection of a representative sample, the objectivity, reliability, and validity concerning the research data plays a significant role (Blasius and Baur, 2014, p. 72).

As described in the previous paragraph, objectivity can be ensured by a large sample and an appropriate research method, such as an online survey, in this case. Research validity and reliability can be improved by applying the concept of triangulation (Flick, 2011, p. 16), which refers to "the combination of methodologies in the study of the same phenomena" (cit. in Flick, 2011, p. 13). Therefore, the questionnaire for this thesis was designed to include qualitative and quantitative questions, which are intended to supplement each other (Blasius and Baur, 2014, p. 42). Whereas the qualitative questions are used to describe, interpret, and comprehend interdependencies, the quantitative questions provide numbers and values aimed at quantifying complex issues in an objective manner (Buber and Holzmüller, 2009, p. 73).

### 3.3. Research Measurement

The author Ian Brace (2013) states that responses to quantitative questions are generally measured using four kinds of data types and associated scales:

- Nominal data, which can be classified "into discrete categories by name" (p. 48)
- Ordinal data, which represents ranking scales without measuring the interval between each rank (p. 49)
- Interval data, which "provide[s] a rating for each item on a scale that has a numerically equal distance between each point" (p. 51)
- Ratio data, which is a special type of interval scale because the zero point has real meaning (p. 53)

This study employs nominal scales, ordinal scales (e.g., in the form of a Likert scale ranging from 1 to 5), and interval scales (e.g., a 10-point rating scale).

However, qualitative questions are assessed and evaluated through the concept of inductive content analysis. Inductive content analysis is recommended when the level of knowledge about a phenomenon is rather low or limited. Therefore, an inductive approach is more appropriate for studies intended to derive a theory from raw data rather than

to verify an existing theory (Elo and Kyngäs, 2008, p. 107). The process of analysis of this method consists of three main stages: preparation, organization, and reporting (cf. Figure 13). However, the main function of content analysis is to classify as many words as possible from a text and to summarize the results in a proper model, conceptual map, or categories (Elo and Kyngäs, 2008, p. 107).

### 3.4. Sample

Before a survey can be developed and distributed, a representative sample has to be defined (Brace, 2013, p. 7). For this empirical study, it was quite difficult to select a representative sample because there is no transparent information available on how many companies accept Bitcoin payments.

According to the merchant service provider Coinbase (2018, 12. February), globally, over 48,000 merchants use its payment processor service. In contrast, the Bitcoin acceptance map called "coinmap.org" displays 11,400 globally registered venues at the time of writing this thesis (Coinmap, 2017, 21. September). A more specific source is btc-echo.de, which gives information about 250 companies accepting Bitcoin in Germany, Switzerland, and Austria and, moreover, there are about 30 multinationals accepting Bitcoin payments (BTC Echo, 2017).

In consideration of the aim of the research questions and the general lack of research available on this topic, this empirical analysis had no specific limitations – the only requirement was that the company had to be registered in Europe. There were no limitations on any particular industry, geographical area, or source of Bitcoin-accepting companies mentioned in the previous paragraph. This implies that this thesis focuses on gathering as much information as possible that can be justified by the research aim of developing a general and holistic framework. Based on these assumptions, the author created a random sample in the form of an Excel spreadsheet including the addresses and corresponding contact person (if any).

The survey was sent to 346 companies/institutions registered in European countries, with the majority (77.72%) being registered in Switzerland, Germany, and Austria. The type of organization ranged from micro-businesses to large corporations. The online survey was initially developed in English and then translated into German. When applicable, the cover letter was personalized, else the survey was sent with a general cover note (cf. Appendix 4).

From November 22 to December 15, 2017, the online survey was fully completed by 65 companies, which is a response rate of approx. 18.79%. An overview of the participants' industries can be seen in Figure 14.

The majority of the survey participants belonged to the retail industry (32.31%), the tourism sector (13.85%), and IT (13.85%). Furthermore, a considerable number operated in the areas of legal, tax, finance, or controlling services (9.23%), other services (7.69%) and in the healthcare industry (6.15%). The remaining quarter was represented by other industries, with a more detailed listing shown in Figure 14.

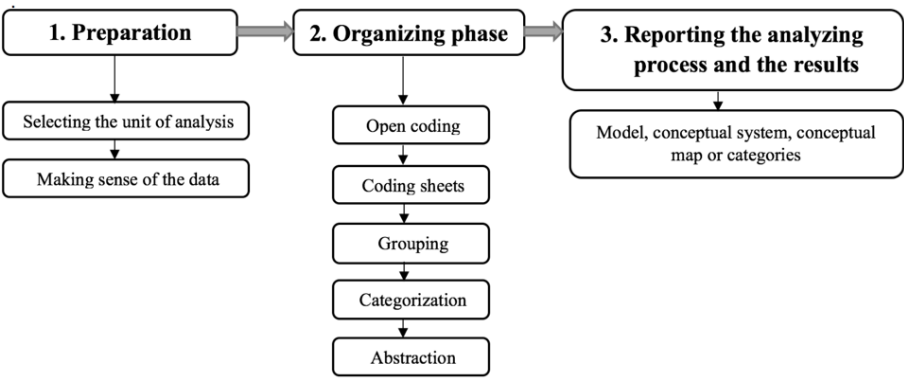


Figure 13: Inductive Content Analysis; Source: Own illustration according to Elo and Kyngäs (2008, p. 110)

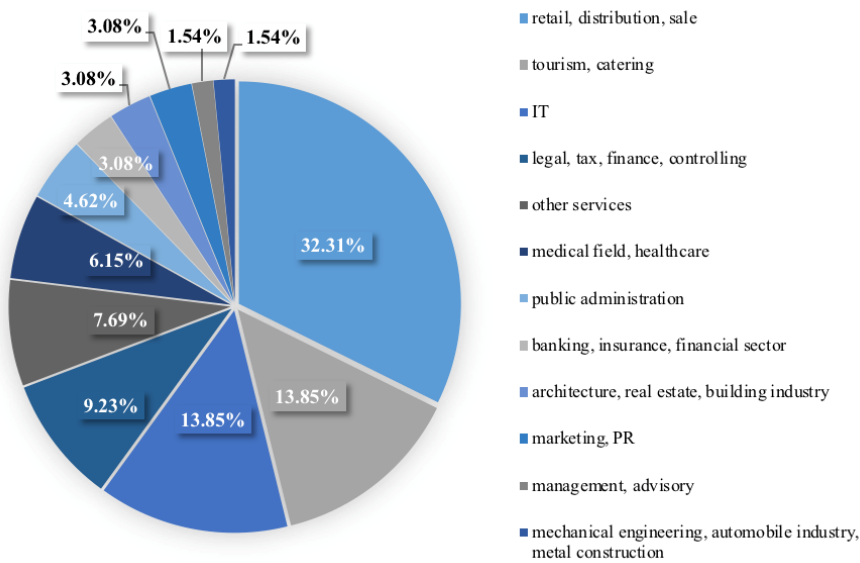


Figure 14: Industry-wide Distribution of Survey Participants

The participating organizations mainly ranged from micro- to small-businesses, with:

- 86.15% having up to 9 employees and up to € 2 million revenue/year (micro-business)
- 9.23% having up to 49 employees and up to € 10 million revenue/year (small-business)
- 3.08% having up to 249 employees and up to € 50 million revenue/year (SME)
- 1.54% having more than 249 employees and over € 50 million revenue/year (large corporation)

With regard to the share of Bitcoin payments (measured as the percentage of total amount of incoming payments), the vast majority (78.46%) indicated that less than 1% of incoming funds are currently paid in bitcoins. However, 18.46% of the participants stated that their share of Bitcoin payments is 1-5% while 3.08% stated the percentage is 5-20%.

4. Methodology: Results

4.1. Risk Identification of Bitcoin

According to Arwinge (2012, p. 45) and the COSO IC Framework (Section 2.2.2.2), risks can be classified into the following three categories:

- Operational risks related to the effective and efficient use of resources
- Reporting risks that might affect financial reporting quality
- Compliance risks representing any breaches of applicable laws and regulations

The identification of risks serves as the basis of empirical risk assessment and, thus, helps answer the research question related to how risks should be evaluated, managed, and controlled by an organization. Therefore, the author has identified and summarized the risks of Bitcoin that might negatively affect an entity's operational, reporting and compliance objectives in order to answer the first research question

and, moreover, to provide the basis for the empirical analysis in Chapter 4.2. Additionally, the potential implications (PI) and potential control activities (PCA) are outlined at the end of each sub-section in the form of a table.

- What are the potential risks and resulting threats of accepting cryptocurrencies as a payment method?

#### 4.1.1. Operational Risks

##### *A lack of know-how*

Cryptocurrencies have emerged with the completely new invention of blockchain technology. Due to the properties of this technology and the way in which it is used to conduct transactions, there are considerable differences in how payments are dealt with compared with other payment methods.

In general, a lack of know-how aggravates all emerging vulnerabilities and risks associated, which will be presented in the following paragraphs. In particular, a lack of know-how with regard to the administration of cryptocurrencies, the purpose of the public and the corresponding private key, and the concept and properties of cryptocurrencies might have implications ranging from reduced operational efficiency to a total loss of all bitcoins that a company owned (provided the company uses only the hot storage option limited to one wallet). In this context, it has to be emphasized that an executed transaction is irreversible, so there is no possibility to recover these lost bitcoins.

A solid education and training on cryptocurrencies might aid the effective prevention and mitigation of these vulnerabilities as well as the associated risks. It is not a prerequisite for a company and its employees to completely understand blockchain technology in order to implement the acceptance of Bitcoin payments. However, a basic understanding of the technology and the necessary risk awareness/evaluation is required. To sum up, a lack of knowledge could have a vast (negative) impact on a company's effective and efficient resource use.

PI: Total loss of all bitcoins, misuse of superior knowledge to commit (irreversible) fraud due to internal or external circumstances, operational losses/ inefficiencies

PCA: Employee training, periodical updates on crypto-markets and price developments, manuals

##### *The Administration of Cryptocurrencies*

As previously described, bitcoins are digital files "with a cash value in the hands of the individual in possession of the file" (Grant and Hogan, 2015, p. 32). If this digital value were to be stolen, the company would have no further opportunities to access its initial bitcoins, and they would finally be lost as there is no way to easily trace back the transaction and reverse it. This is comparable to losing one's purse and all the cash inside it (Choo, 2015, p. 300). Particularly with regard to the dealing and storage of this digital value, an organization's management must address two interrelated questions:

first, who is responsible for the (private) key administration and, second, how is this key to be stored?

The owner of the private key has the full access rights and ownership of the corresponding bitcoins, which implies that safekeeping is the most important role in the operational dealing of cryptocurrencies. In a risk report, the insurance company Lloyds (2015, p. 7) described the private key as a digitized "secret such as banking credentials, intellectual property, or private photographs," which could be or could become a decisive factor in an organization's success.

Therefore, it plays an important role in not only the administration of this key within a company but also in clearly determining the responsibilities and duties concerning the administration of the bitcoins received. In this context, there is also the possibility to outsource the private key's administration to an intermediary or an external provider like a payment processor or an exchange platform (Lloyds, 2015, p. 7). Key storage by an external provider offers the advantages of a professional and secure administration (Franco, 2014, p. 131; Grant and Hogan, 2015, p. 33). However, this is also related with additional costs and risks, especially with regard to the reliability of an external service provider. History has already unveiled such risks; for instance, the exchange platform Mt. Gox was hacked in 2011 and 2014 which resulted in around 650,000 bitcoins being stolen from its customers and the company finally suffering insolvency (Hern, 2014, 18. March).

Regardless of whether storage is outsourced or not, there are in general two types of wallets used to keep bitcoins: hot and cold storage wallets. Blockchain researcher Toshendra Sharma (2017, 14 November) states that a hot wallet can be compared with a checking account, whereas a cold wallet is a kind of savings account. The question then arises of which type of wallet a company chooses and how these wallets will be secured. Although a cold wallet is an offline wallet that is not connected to the Internet, hot wallets or web wallets are stored on devices that are connected to the Internet and directly communicate with the Bitcoin network (Franco, 2014, p. 126).

This cold storage option is available in several forms: the private key could be saved on an external media storage (e.g., USB flash drives or optical disks), on a paper wallet (by printing the private key on a piece of paper), or on a special hardware wallet like Trezor (Antonopoulos, 2015, p. 237; Franco, 2014, pp. 126f.). The benefit of a cold wallet is that it is secure against hacker attacks, but due to their physical and tangible form, they require appropriate storage facilities like a vault (Franco, 2014, pp. 17-18). In contrast, a web-based wallet is usually provided by an external provider and is in the form of an online account (with additional features) in which funds can be deposited and accessed, similar to online banking (Franco, 2014, p. 131). Users can access the account by means of an authentication process (e.g., 2-factor authentication or multi-signature), after which they can conduct transactions. Additionally, the use of a hot wallet is more appropriate in order to accept payments because of a faster verification process (Sharma, 2017, 14 November).



Depending on the choice of wallet type and/or third-party involvement, the safekeeping of keys requires appropriate cyber security measures, physical protection, or both. In general, it is recommended to store larger amounts of cryptocurrency assets in cold wallets, while hot wallets would be a better choice for daily operations. (Antonopoulos, 2015, p. 237; Sharma, 2017, 14 November)

PI: Total loss of bitcoins due to inappropriate hot storage, physical theft due to unreasonable cold storage, employee fraud in operational dealings  
PCA: Outsourcing of "cold" private key storage, e.g., in the vaults of multinational banks, limitations of access rights to the company's wallet, multi-signature wallets, encryption of private keys, diversification of bitcoins (hot and cold wallets), implementing holding limits on the hot wallets

#### IT Risks

Particularly in the context of IT, recent technological innovations have led to the emergence of new risks. The two kinds of risks associated with Bitcoin are global and local. Global risks or "attacks" are aimed at manipulating the entire Bitcoin network, while local risks or "attacks" are characterized by the intention to get specific access to a private key or password of a company's hot wallet "in order to gain control of bitcoins and the matched public address" (Lloyds, 2015, p. 7). On account of significant price developments and the resulting increase in the value of bitcoins, the cryptocurrency is an extremely attractive target for hackers and thus for local attacks. This is further encouraged by the instant diversion and irrevocability of transactions without the need of converting "identity information or account tokens – such as credit cards, and bank accounts – into value after compromising them" (Antonopoulos, 2015, p. 236).

Therefore, vendors accepting Bitcoin should be aware of the importance of private keys (or wallet passwords). For instance, they should not store this key in the form of any other data file (including encrypted files) on their desktop computers (Antonopoulos, 2015, p. 236). However, the risk of a local attack can be mitigated by means of appropriate cyber security measures, the necessary know-how, and the appropriate distribution of coins on different storage mediums.

Controversially, global risks (e.g., a 51%-attack or a Distributed-Denial-of-Service (DDoS)-attack) which are closely related to some vulnerabilities of blockchain technology, cannot usually be controlled and influenced by the company itself (Lloyds, 2015, pp. 10f.). The scope of this thesis, however, does not cover an in-depth analysis of these IT-related (blockchain) risks.

PI: Private key/ password theft (depending on whether hot or cold storage) through hacking, skimming, etc., and the resulting operational losses

PCA: Appropriate IT infrastructure/ risk management, implementation of additional encryption standards, several storage options (hot and cold storage)

#### The Irreversibility of Transactions

In general, the irreversibility of transactions can be seen as an advantage but also as a source of some risks. For instance, if a local attack were to be successful and the "thief" were to gain access to a company's private key, the bitcoins could be grasped immediately and distributed to different wallet addresses (Lloyds, 2015, p. 8). In such a case, there would be no possibility of reversing the transactions, i.e., there is no charge-back option such as in the case of a disputed service by a credit card provider (IMF, 2016, p. 29; Vigna and Casey, 2016, p. 127). Moreover, one has to keep in mind that a customer has no right to refund a transaction. If anything goes wrong, the risks of default payments have to be covered by the Bitcoin network participant or the Bitcoin accepting vendor, because there is no central authority that has the responsibility to do so (IMF, 2017, p. 29). This makes a fraud case especially challenging since there is no relevant authority that could freeze or seize an order (IMF, 2017, p. 28).

However, this irreversibility can also be an advantage given the new digital business models, e.g., when digital files are instantly delivered and cannot be reversed, like in the case of the download of songs or software (Sixt, 2017, p. 89).

PI: Operational losses due to default payments, wrong transactions, and fraud  
PCA: n/a

#### 4.1.2. Reporting Risks

##### Reporting Standards

Grant and Hogan (2015, pp. 30-31) argue that the explicit reporting and classification of the bitcoins received depends on their usage, i.e., either as a medium of exchange or as an investment. Nevertheless, there is not yet a consistent regulation concerning the explicit classification of bitcoins (Lindner and Meyer, 2017, p. 29). In the following paragraphs, the reporting of bitcoins will be analyzed in the context of Germany and Switzerland.

Lindner and Meyer (2017, p. 29) have questioned the explicit reporting of bitcoins as "liquid funds" in Switzerland, given their significant price volatility; however, it might be also possible if this requirement were to be fulfilled. According to Heer, 2014, 10. December, the classification of bitcoins as liquid funds is excluded. Instead, bitcoins should be reported as "short-term assets with market price" and thus should be classified as short- or long-term securities (Heer, 2014, 10. December; Lindner and Meyer, 2017, p. 29). In summary, the explicit classification depends on the usage of bitcoins, for instance, if the bitcoins are used for investment purposes and will be held for a longer term, the coins should



be classified as long-term financial assets such as commodities like gold or oil (Lindner and Meyer, 2017, p. 29).

In Germany, the analysis of Kanton (2017, p. 2735) reveals that according to the German Handelsgesetzbuch, bitcoins have to be reported as "other assets," part of the current assets, or as "acquired intangible assets" as part of the fixed assets. It is extremely important to comply with the national reporting standards, because a violation could lead to the fraudulent misrepresentation of the financial statements, an inaccurate company valuation, and issues with tax evasion.

PI: (Fraudulent) misrepresentation of financial statements, inaccurate company valuation, evasion of tax

PCA: Establishing guidelines for the national reporting of Bitcoin

#### *The Price Volatility & Exchange Rate Risks*

One of the most common risks with cryptocurrencies and, Bitcoin especially, is the considerable degree of price volatility. Appendix 2 contains an extracted chart of the Bitcoin/USD spot rate from January 16, 2017, to January 16, 2018. It shows the significant volatility of the cryptocurrency. Whereas one bitcoin was traded at around USD 800 at the beginning of 2017, the value increased by 1400% to approx. USD 12,000 in January 2018. In order to further exhibit these price fluctuations, the Bloomberg "Return-on-Change" function (the yellow line below the price index chart) was added. This function is a technical indicator of volatility, measuring the price change between current price and the price ten days ago in percent. Therein, substantial fluctuations of 20% are no rarity.

The extreme price volatility of incoming funds first has an impact on the exchange rate risk (FX risk), i.e., the price at which the bitcoins will be converted into national currencies. In turn, this might affect the entity's effective and efficient operational use of resources as well as the company's financial reporting.

With regard to operational effectiveness, it can be primarily stated that the received funds could appreciate but also depreciate within in ten days by about 15-20% (cf. Appendix 2), and thus significantly affect the operational revenue within a very short period of time. In addition, assuming that the company uses Bitcoin funds for operational purposes and as a medium of exchange, it could even lead to liquidity shortages or considerable FX losses, for instance, in case of paying a supplier if the price drops by about 20% in 72 hours (Grant and Hogan, 2015, p. 30).

Financial reporting might also be influenced by these FX fluctuations. On the one hand, it is quite difficult to match any revenue forecasts or conduct appropriate budgeting (Hoelscher, 2014, p. 24). On the other hand, depending on the usage of Bitcoin and the applicable national law, the exchange of bitcoins in national currencies could result in tax charges.

PI: Operational losses due to FX fluctuations, liquidity shortages, tax charges, inability to convert Bitcoin into other currencies due to a lack of liquidity

PCA: Hedging of bitcoins, instant conversion by means of a payment processor, establishing guidelines related to the monitoring of liquidity in Bitcoin markets

#### 4.1.3. Compliance Risks

##### *Tax-Compliance*

In general, more is known about the taxation of Bitcoin than about its financial reporting (Lindner and Meyer, 2017, p. 28). Nevertheless, cryptocurrencies exhibit a high potential for tax evasion due to the fact that participants do not need to disclose their identity, and there is no means to identify beneficial owners (IMF, 2016, p. 30). For appropriate tax treatment, it is first important to know who is the subject of the tax, i.e., a natural person or a judicial body. Additionally, there is also the question of the classification of bitcoins, i.e., whether it is a form of currency or a short- or long-term security.

In case of a judicial body, the received payments have to be recognized in the income statement (calculated in the national currency by the time the funds are received) and could be further activated in the financial statements (see Section 4.1.2). If they are instantly converted to fiat currencies, they will be recognized as revenue, and there would be no further tax issues. Controversially, if the bitcoins are held as short- or long-term security, the resulting capital gains (in case of realization) would be additionally subject to the corporate income tax in Switzerland or the personal income tax and corporate income tax in Germany (Lindner and Meyer, 2017, p. 29; Schmidt, 2018, 18. January).

In Switzerland, a natural person is not faced with any specific tax-compliance issues if the bitcoins are used for private purposes, and even if any capital gains are realized with exception of Swiss property tax (Kanton, 2017).

In contrast, in Germany, any bitcoin capital gains are subject to German income tax regulations. Only if a natural person holds the bitcoins for a period longer than one year, are the achieved capital gains tax free (Schmidt, 2018, 18. January). The relevant valuation method is the first-in-first-out method, which can be made transparent by means of the blockchain (Schmidt, 2018, 18. January).

PI: Tax evasion and its legal implications

PCA: Establish guidelines for the recognition and valuation of Bitcoin transactions with an emphasis on regulatory compliance

##### *Anti-Money-Laundering and Know-Your-Customer Issues*

From a financial integrity perspective, the properties of Bitcoin do raise concerns, particularly with regard to its anonymity and the cross-border reach of transactions (IMF, 2016, p. 27). Although every transaction is publicly available, the users are virtually anonymous (Grant and Hogan,

2015, p. 32). This implies that the concept of Bitcoin has a high degree of transparency in the transactions conducted but not with regard to who has conducted the transaction. This is sometimes described as pseudo-anonymous, because of which cryptocurrencies are fundamentally vulnerable to money laundering. The IMF (2017, p. 27) states that bitcoins can be used "to conceal or disguise the illicit origin or sanctioned destination of funds, thus facilitating money laundering (ML), terrorist financing (TF), and the evasion of sanctions." Furthermore, bitcoins are likely to be used for cyber-related crimes, e.g., on the "dark web" or as ransom money to decrypt data previously encrypted by malware.

Regulators have adopted different approaches to mitigate these risks (cf. Chapter 2.1.4). In general, only financial intermediaries are faced with AML- and KYC-related regulations, so the vendors and their users have not yet faced these compliance issues (IMF, 2016, p. 28; Pillar, 2017, pp. 8-9; Choo, 2015, p. 304). This implies that a vendor accepting Bitcoin has to take appropriate measures to ensure that his customers are in line with the KYC- and AML- related laws.

PI: Accepting illicit funds, being liable to prosecution

PCA: Customer due diligence, reporting of suspicious transactions

## 4.2. Empirical Analysis

The following empirical analysis confronts the participating organizations with the risks identified in order to answer the second research question:

- Are the companies and the responsible employees aware of these risks, and how do they evaluate and address them?

### 4.2.1. Payment Process

If a company accepts Bitcoin payments, it is first necessary to understand how it receives and deals with them, i.e., to what extent the transaction process is self-managed and for which purposes the received bitcoins are used.

The analysis reveals that 44.62% of the participants employ a merchant service provider to process the incoming Bitcoin funds, whereas 55.38% indicated that they manage the payments without any third-party involvement (cf. Figure 15). Since the use of a payment processor can help mitigate or eliminate some emerging risk factors (e.g., the FX risk or the payment default risk), the analysis made a distinction between participants who do and do not use merchant service providers.

We found that around half of the participants in our survey use payment processors. Although there are many reasons behind this decision, Figure 16 enable the derivation of two overriding reasons.

Note: The participants were allowed to choose multiple answers and the sum of the results is therefore over 100%.

Figure 16: Reasons for Using a Payment Processor The first reason is the mitigation of risks. Among the respondents, 62.07% mentioned using payment processors for the

prevention of the FX risk through the possibility of instantly exchanging bitcoins into fiat currencies, and 44.83% cited covering a potential payment default due to the immediate payment confirmation and the liability of the merchant service provider. Moreover, the appropriate documentation and invoicing that is closely related to appropriate financial reporting and the resulting reporting risks, was mentioned by 31.03% of the participants.

The second dominant reason for involving a merchant service provider was for simplification and convenience, which leads to lower entry barriers for potential vendors accepting Bitcoin payments. In fact, 58.62% of the respondents justified using a merchant service provider because of the infrastructure provided, such as a shopping cart plugin-in for e-commerce or applications and interfaces for point-of-sales. Furthermore, even 58.62% of the participants cited implementing a payment processor because it enables the acceptance of Bitcoin payments without in-depth knowledge.

The analysis also investigated the other half of participants who do not use third-party providers because of its potential redundancy and the costs involved. Therefore, the analysis focused on the awareness of risks that might be potentially covered or mitigated by means of payment processor, i.e., the FX risk, defaulting payments and the proper documenting and invoicing which is closely related to reporting and tax risks.

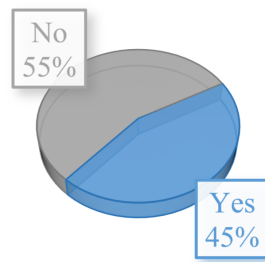
The results show that 80% neither experienced a payment default nor viewed the documentation and invoicing as a pitfall, which is consistent with the initial assumption that a payment processor would be redundant. Controversially, the FX risk which could be almost completely eliminated, was mentioned by 45.71% of the participants as a risk, and 37.14% recognized the reporting and tax issues as a risk.

This study also addressed the question of how a company deals the incoming Bitcoin funds. In general, there are two different possibilities: either bitcoins are only used as a payment method and they will be instantly exchanged into national currencies, or the bitcoins will be held by the company for use as liquid funds or as an investment. In this context, storage plays a critical role, because if the bitcoins are used as an investment or as a liquid fund in order to pay invoices or suppliers, the company needs an appropriate Bitcoin administration. Controversially, if the bitcoins are exchanged immediately or on the following day, the Bitcoin storage becomes more or less obsolete.

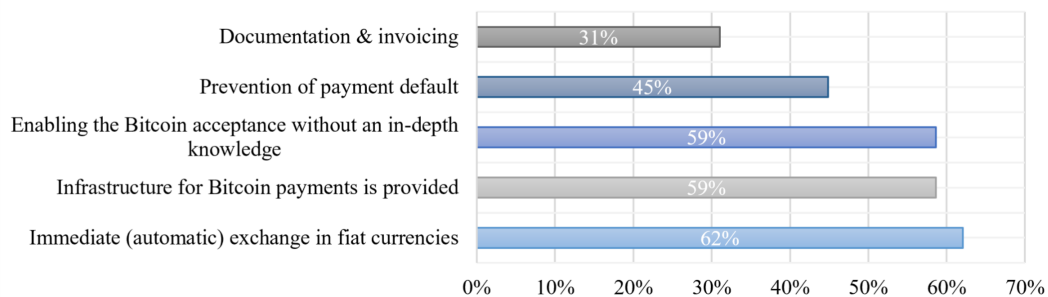
The pie chart in Figure 17 shows the results concerning the handling of incoming Bitcoin payments. It is apparent that almost 62% (using cryptocurrency as liquid funds or investment) of the respondents store the incoming funds, whereas around 38% exchange incoming funds instantly or by the following day.

The 62% of participants storing the funds are especially faced with the threat of proper Bitcoin administration and all its related risks. Hence, the second part of the empirical investigation analyzes the administration and duties related to Bitcoin payments.

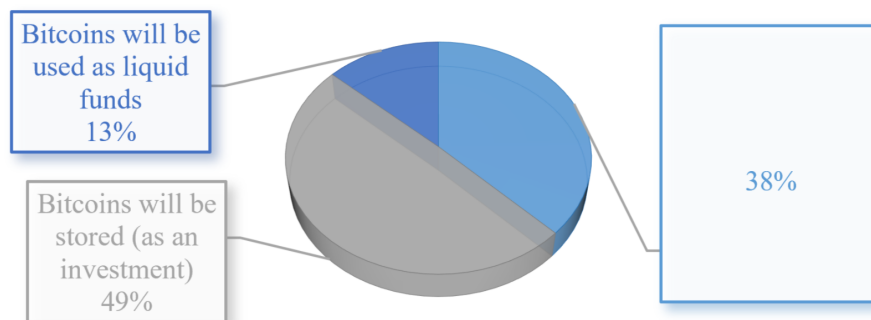
In summary, the first part of this evaluation concerning



**Figure 15:** Third-Party Involvement in the Payment Process



**Figure 16:** Reasons for Using a Payment Processor; Note: The participants were allowed to choose multiple answers and the sum of the results is therefore over 100%.



**Figure 17:** Dealing with Incoming Bitcoin Funds

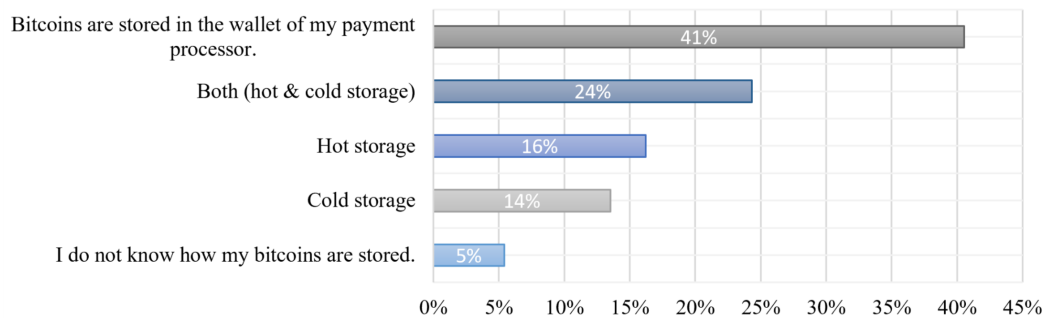
the payments process reveals that there is no definite trend of how this process plays out in practice, i.e., with or without a merchant service provider. This implies a considerable issue for the development of a proper Cryptocurrency IC Framework, because all potential risks have to be addressed. Furthermore, the results of the present study demonstrate that the use of a payment process is justified by two reasons: risk mitigation/elimination and the purposes of simplification and convenience. In context of risk mitigation, the FX risk was mentioned as the greatest danger which could be basically eliminated. Moreover, approx. two thirds of the companies store the incoming funds and are thus confronted with a reasonable administrations of Bitcoin payments.

#### 4.2.2. Administration of Bitcoin Payments

It is important to understand that there is not necessarily a connection between Bitcoin storage and the involvement of a payment processor. Even though a kind of third-party provider is employed, cryptocurrencies might be transferred and stored in one's "own" wallets. Therefore, the crucial question is if the bitcoins are stored or not, whether in the form of an investment or as liquid funds, as discussed in the previous chapter.

When bitcoins are stored, a company is faced with several options that are listed in Figure 18.

As can be seen in Figure 18, the analysis indicates that the majority of participants (40.54%) store their bitcoins in a merchant service provider's wallet. In this context, the private key storage is outsourced, with the resulting risk of third-party reliability. When only the hot storage method is used



**Figure 18:** Options for Bitcoin Storage

(16.22%), the company faces the same advantages, but the difference is that the risk of third-party reliability is due to the wallet provider and not the payment processor. In both cases, the administration of rights to access the third-party wallet plays a significant role.

Moreover, we found that 24.32% of companies use both methods, and 13.51% only use the cold storage option. Therefore, the storage and administration have to be clearly determined. Finally, 5.41% of the respondents stated that they did not know how their bitcoins are stored.

In addition to Bitcoin storage, this study also investigated who is generally responsible for the administration of bitcoins, i.e., who has the right to access the private key or wallet and to which field of competency this duty is assigned. Several noteworthy results in this context were the answers "nobody" and "everybody" in response to the question, "Who has the admission or the relevant rights to access your company wallet?". With the exception of these answers, the majority of participants (90.91%) mentioned that a specific person has access to the company's wallet. The responsible persons are typically the Chief Executive Officer (CEO), Chief Digital Officer (CDO), IT Administrator or Head of Accounting, as well as their assistants or representatives.

Responses to the question regarding the number of people who have the access rights ranged from one to ten employees, with an average of two persons having the rights to access the wallet. With regard to their field of competency, this duty is typically assigned to the senior management (CEO, CDO, Chief Technology Officer (CTO) or the Chief Financial Officer (CFO)), or an executive manager such as the IT Administrator or the Head of Accounting.

#### 4.2.3. Risk Awareness & Evaluation

The third part of the empirical analysis focused on the awareness and evaluation of the risks identified in Chapter 4.1.

The respondents were first asked to describe their awareness of the risks by evaluating if there is a risk potential in general and how likely it is that this risk will occur. It was possible to choose from five options ranging from "no risk" to "extremely high risk." The results are summarized in Figure 19 below.

The data suggests that the experienced potential of the risks identified is rather low. In particular, the irreversibility of the transactions does not pose a risk for 40.32% of the participants, whereas 27.42% rated it as a low risk, and 20.97% as a moderate risk. This implies that only around 5% considered the irreversibility of the transactions to be a threat. Furthermore, risks related to the financial reporting and taxes were rated comparatively low. In contrast, the price volatility and the resulting FX risk of Bitcoin exhibited the highest risk potential, with 23.44% of the participants classifying it as an extremely high risk and 18.75% as a high or moderate risk. Moreover, a lack of know-how, risks within the administration of Bitcoin, and cyber risks were ranked as moderate risk or greater by at least 45% of the participants.

Besides the risk evaluation, the results shown in Figure 19 also reveal that 9.52% of the participants were not able to assess the compliance risks. With regard to the blockchain risks (7.81%), the cyber risk (6.35%), the financial reporting risk (6.25%), and the irreversibility of transactions (6.45%), an increased uncertainty related to the evaluation of the risk potential was also apparent.

Nevertheless, in order to clearly interpret the results, the authors calculated the mean of the risks evaluated (no risk – 1, extremely high risk – 5) based on the responses. Subsequently, the risks were classified and ranked according to their classification and risk potential, which can be seen in Table 3.

The analysis also aimed at investigating the severity of the potential of damage that could be caused by the risks identified. Here, the participants had the option to evaluate the risk severity by choosing a number between one and ten in order to assess and further express the damage potential from their point of view. Here, one represented a low or almost no damage potential, and ten an extremely high damage potential. The results can be seen in Table 4.

The figures reveal that the companies generally assess the severity of potential damages as low. Their greatest concerns are about risks related to FX, cyber, and blockchain. Controversially, financial reporting, compliance, tax risks, and the irreversibility of transactions are considered to have a lower potential for damage.

These results have been summarized in a risk matrix cre-

**Table 3:** Evaluation of the Risk Potential

|                        | Price volatility<br>& FX risk<br>of Bitcoins | A lack of<br>know-how | Cyber<br>Risks    | No clear responsibilities<br>concerning the<br>Bitcoin administration | Blockchain<br>risks | Compliance<br>risks | Tax<br>risks | Financial<br>reporting<br>risks | Irreversibility<br>of transactions |
|------------------------|----------------------------------------------|-----------------------|-------------------|-----------------------------------------------------------------------|---------------------|---------------------|--------------|---------------------------------|------------------------------------|
| Mean                   | 3,14                                         | 2,66                  | 2,57              | 2,56                                                                  | 2,29                | 2,28                | 2,18         | 2,03                            | 1,91                               |
| Risk<br>Classification | moderate                                     | low -<br>moderate     | low -<br>moderate | low -<br>moderate                                                     | low                 | low                 | low          | low                             | low                                |



Table 4: Evaluation of the Risk Severity

|                                     | Price volatility<br>& FX risk<br>of Bitcoins | Cyber<br>Risks    | Blockchain<br>risks | A lack of<br>know-how | No clear responsibilities<br>concerning the<br>Bitcoin administration | Financial<br>reporting<br>risks | Compliance<br>risks | Tax<br>risks | Irreversibility<br>of transactions |
|-------------------------------------|----------------------------------------------|-------------------|---------------------|-----------------------|-----------------------------------------------------------------------|---------------------------------|---------------------|--------------|------------------------------------|
| Mean                                | 3,66                                         | 3,58              | 3,37                | 3,08                  | 3,04                                                                  | 2,96                            | 2,95                | 2,7          | 2,02                               |
| Risk Severity<br>(damage potential) | low -<br>moderate                            | low -<br>moderate | low -<br>moderate   | low -<br>moderate     | low -<br>moderate                                                     | low                             | low                 | low          | low                                |

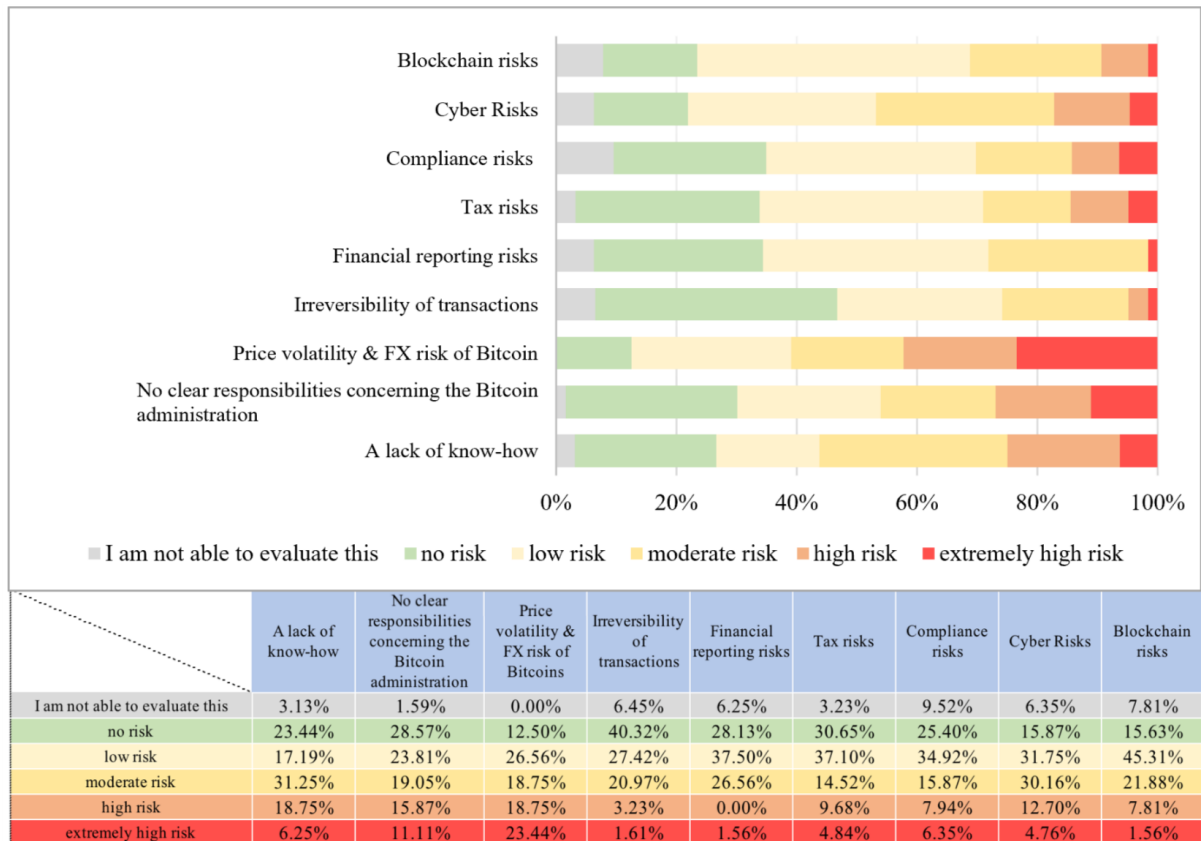


Figure 19: Evaluation of the Risks Identified

ated by the author based on the data available. Taking into account the centered extract (cf. Figure 20) of the established risk matrix (the complete risk matrix can be found in Appendix 6), we can conclude that the FX risk is experienced as the greatest threat. Cyber and blockchain risks as well as the lack of responsibility concerning know-how and Bitcoin administration play a comparatively significant role. In particular, the irreversibility of the transactions was ranked as a lower risk.

#### 4.2.4. Risk Handling

The last part of the empirical study deals with the potential and practical handling of the risks identified. We considered only those risks that can be effectively influenced by a company, for instance, a company cannot influence price volatility or blockchain risks but it can control the exchange rate or cyber risks.

The potential development of internal controls was analyzed in this regard. According to the results, especially regarding insufficient know-how and the segregation of duties (no clear responsibilities concerning Bitcoin administration), the development of internal controls seems possible. Around 70% of the respondents stated that the development is rather or absolutely possible. However, only 30% were of the opinion that it is rather not possible respectively impossible to develop internal controls in order to hedge the FX risk of Bit-

coin. A detailed overview of these responses can be seen Figure 21.

In addition to the potential development, the practical implementation was also analyzed. In this context, only 46.15% of the companies have implemented internal controls or similar measures in order to ensure the appropriate dealing of Bitcoin payments (cf. Figure 22).

In this context, the explicit dealing was scrutinized, i.e., which specific internal controls or measures have been implemented. The results are summarized in Appendix 7 and serve as an important component of the risk control matrix provided in Chapter 4.3.1.

The results indicate that the risks identified are properly addressed only by a few companies. In order to overcome a lack of know-how, several organizations have implemented regular training sessions or workshops on cryptocurrencies, their purpose, and dealing. Some companies also provide their employees with manuals and have created internal guidelines intended to impart the knowledge required for operational business. The staff also gets the thus so far voluntary opportunity to watch YouTube tutorials during work hours as a form of e-learning, so as to increase their level of knowledge.

With regard to the appropriate administration and storage of bitcoins, many enterprises have decided to limit the responsibilities to an extremely small group of people, for in-

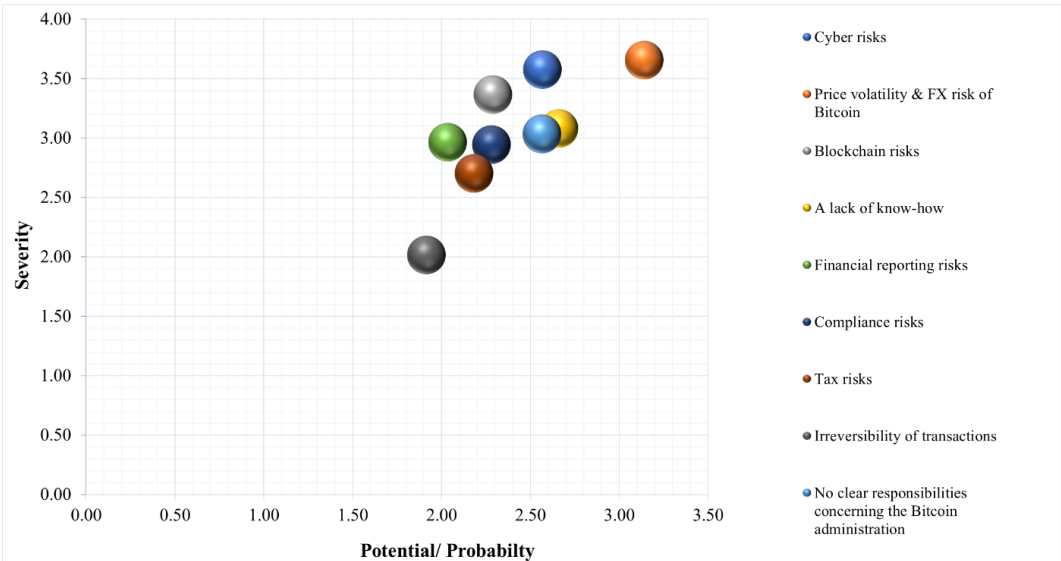


Figure 20: Extract of the Risk Matrix

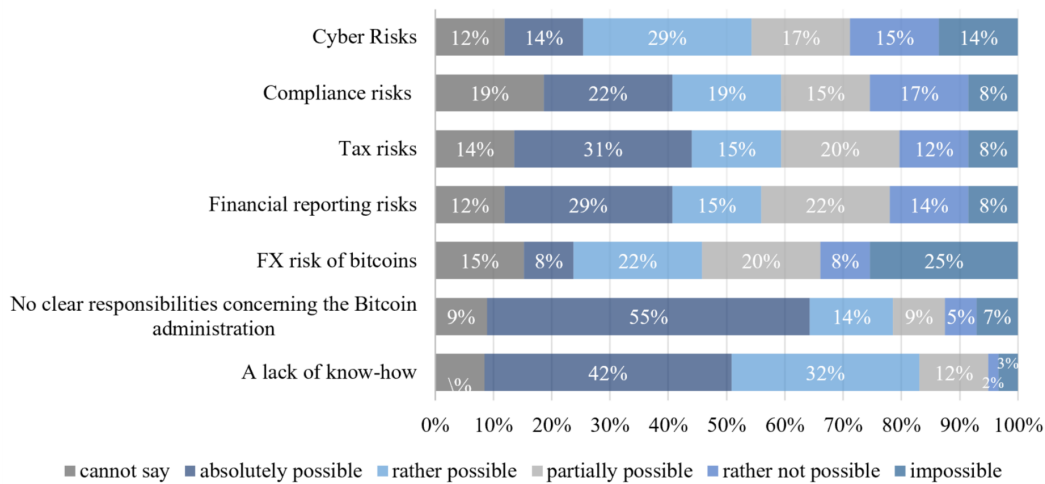


Figure 21: Development of Potential Internal Controls

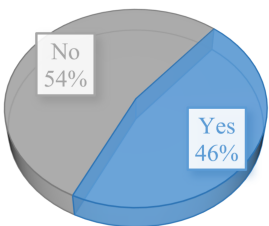


Figure 22: Implementation of Internal Controls

stance, only the CEO and his or her representative typically have rights to access the company’s wallet. In this context, several companies emphasized that the responsibilities have to be clearly documented and executed in order to prevent a segregation of duties. The storage options have already been explicitly analyzed in the previous chapters; however,

a few companies were of the opinion that the cold storage itself represents a form of control measure. Additionally, one company is using a software that encrypts the corresponding pins/keys to further enhance its level of security. The FX risks and the relevant internal controls are more or less obsolete if an enterprise instantly converts its received

Bitcoin funds into fiat currencies. As previously discussed, most companies indicated that it is impossible or rather not possible to develop internal controls, which is consistent with the findings concerning the internal controls related to FX risks. Besides the opportunity to completely eliminate the exchange rate risk through instant conversion, the participants mentioned that observation, analysis, and relevant documentation are also internal control measures taken to mitigate the FX risk.

In the area of reporting risks – the identified financial reporting risks but risks related to tax-compliance – the implemented IC measures are quite similar. In both cases, the appropriate documentation and the "four-eyes" principle were mentioned as the most important internal controls. Nevertheless, manual control or the additional involvement of an external advisor was proposed as a useful supplementary measure.

Furthermore, the analysis revealed several interesting IC measures regarding risks related to KYC and AML issues. Primarily, it can be stated that some companies use additional KYC checks in form of an obligatory part of the payment respectively ordering process, with suspicious customers being immediately reported to the authorities. The internal mapping of public addresses to real customers/identities in order to get a more transparent customer profile is yet another approach. The involvement of an internal IT Forensic division was mentioned as an internal control as well. Apart from these measures, the permanent monitoring of changes in the national regulations is certainly required to overcome any compliance risks.

The involvement of an internal IT Forensic division (if any) plays a significant role in mitigating cyber risks as well. By means of such a division or through internal IT risk management, potential threats can be prevented or detected in their early stages. Conducting training sessions or workshops to enhance the risk awareness and knowledge of companies' employees is also an appropriate measure. If a company would like to bring down the cyber risks (of Bitcoin storage) to an acceptable level of risk, they could even consider using cold wallets and an additional encryption software.

### 4.3. Recommendations with Regard to Cryptocurrencies

Based on the results of the previous two research questions, the author established a risk control matrix in order to provide a clear overview and summary of the findings. These conclusions will be incorporated into a holistic Cryptocurrency IC Framework, which will be developed along the lines of the COSO 2013 IC Framework. This will finally lead us to the answer of the third research question:

- Are there any general control activities and recommendations necessary to ensure an appropriate risk management in the handling of cryptocurrencies?

#### 4.3.1. Risk Control Matrix

The author has summarized the findings of the previous two research questions in the form of a risk control matrix

to provide comprehensive and transparent overview of the risks identified and to outline recommendations concerning their handling. As proposed in Chapter 2.2.2.2, the matrix is divided into three superordinate categories – objectives, risk analysis and control activities – which will be further specified.

The first category of "objectives" is sub-divided in two sections: the first column shows the general objectives of the risk control matrix with regard to cryptocurrencies, and the second column categorizes the general objectives according to the three classifications proposed by the COSO 2013 IC Framework.

The following three columns represent a subdivision of the category of "risk analysis" and display the findings of the investigation. Therefore, it displays the identified risks (Chapter 4.1) and their corresponding risk potential and severity (cf. Chapter 4.2.3) as evaluated by the survey participants.

Finally, the last category summarizes the "control activities." This was done by selecting the appropriate measures proposed in the empirical analysis in Chapter 4.2.4 and by adding the potential control activities mentioned in Chapter 4.1. Additionally, the author assessed the types of controls (preventive = pr, directive = di, detective = de or corrective = cr) and the forms of the control activities, i.e., automated (a) or manual (m). The entire risk matrix can be found in Table 5.

#### 4.3.2. Cryptocurrency IC Framework

The third research question also aimed at identifying general recommendations to ensure appropriate risk management in the handling of cryptocurrencies. Hence, the author decided to establish a framework that serves as a set of guidelines to implement and handle cryptocurrencies as a payment method and to ensure appropriate risk management.

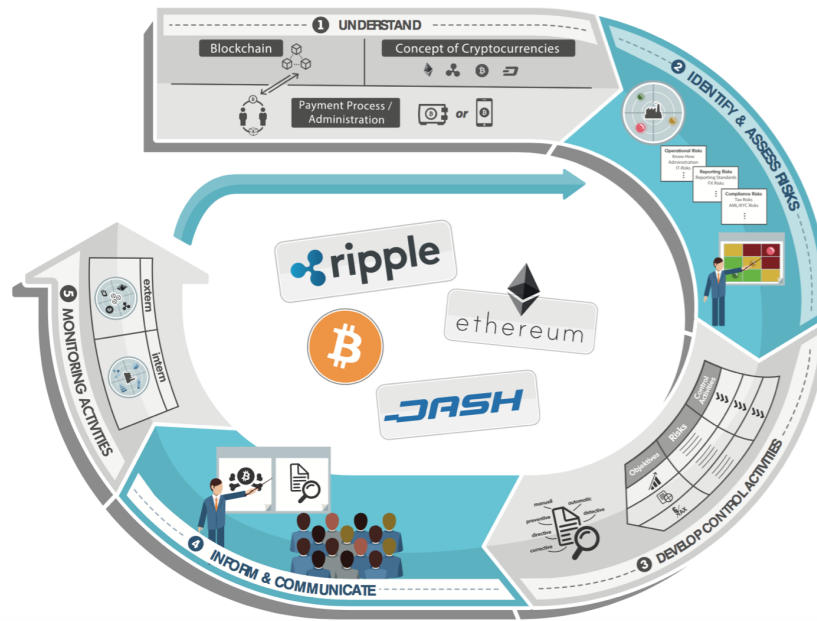
The Cryptocurrency IC Framework developed is along the lines of the COSO 2013 IC Framework and adapted to cryptocurrencies based on the research results obtained through the example of Bitcoin. The framework can be seen Figure 23, with a higher resolution version in Appendix 8. Figure 23 shows a framework based on five processing steps. However, the established Cryptocurrency IC Framework does not present a "5-stages-model" – it should be rather understood as a continuous circle of processes. Moreover, it is important to understand that the framework represents a holistic approach and should thus be viewed as a whole and not just as a collection of its five components, each of which is described in further detail below:

1. Understand: Even though the analysis revealed that some or the other risk is experienced as more or less dangerous, the holistic model does not start by tackling specific risk factors. The fundamental prerequisite for the proper implementation and dealing of cryptocurrencies is a basic knowledge of its technology – the blockchain – the concept of a cryptocurrencies, the payment process, and the corresponding administration of

**Table 5:** Evaluation of the Risk Severity

| Objectives                                                                                                                                                                                                   |                              | Risk Analysis                                                                |                             |                     | Control Activities                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|------------------------------------------------------------------------------|-----------------------------|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| General Objective                                                                                                                                                                                            | Classification of Objectives | Identified Risk                                                              | Risk Potential/ Probability | Risk Severity       | Internal Controls/ Measures                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Type Form                                                                           |
| following the COSO IC Framework                                                                                                                                                                              |                              | (cf. Chapter 4.1)                                                            | (cf. Chapter 4.2.3)         | (cf. Chapter 4.2.3) | (cf. Chapter 4.2.4) + own supplementation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | (own assessment)                                                                    |
|                                                                                                                                                                                                              |                              | A lack of know-how in dealing with cryptocurrencies                          | 2.66                        | 3.08                | <ul style="list-style-type: none"> <li>- establishing manuals and internal guidelines</li> <li>- giving regular employee trainings/ workshops</li> <li>- providing regular market updates</li> <li>- establishing a proper e-learning concept, e.g., YouTube tutorials &amp; quizzes</li> <li>- using multi-signature wallets</li> <li>- outsourcing of private key storage/hardware wallets</li> </ul>                                                                                                                                                              | pr/di<br>m<br>pr/di<br>m<br>pr/di<br>a/m<br>pr/di<br>a<br>pr/di<br>a/m              |
|                                                                                                                                                                                                              |                              | A lack of clear responsibilities in relation with the Bitcoin administration | 2.56                        | 3.04                | <ul style="list-style-type: none"> <li>- limiting of responsibilities to a small group of persons</li> <li>- access rights have to be clearly documented</li> <li>- reading of the manual is required</li> <li>- additional pin encryption, for instance by using the software keepass</li> <li>- using multi-signature wallets</li> <li>- diversification of bitcoins to several wallets (hot &amp; cold wallets)</li> <li>- establishing limits of the bitcoin holding amount on hot wallets and transfer the excess number of bitcoins to cold wallets</li> </ul> | pr/di<br>m<br>pr/di<br>m<br>pr<br>oca<br>pr/di<br>a<br>pr/di<br>a/m<br>pr/di<br>a/m |
|                                                                                                                                                                                                              |                              | Cyber risks                                                                  | 2.57                        | 3.58                | <ul style="list-style-type: none"> <li>- implementing a reasonable IT risk management</li> <li>- establishing an internal IT forensic division</li> <li>- using cold wallets or outsourcing of private key storage</li> <li>- giving regular employee trainings/ workshops</li> <li>- implementing additional pin encryption, e.g., by using the software "Keepass"</li> </ul>                                                                                                                                                                                       | pr/de/cr<br>a/m<br>pr/de/cr<br>a/m<br>pr/di<br>a/m<br>pr/di<br>m                    |
| Development and maintenance of an internal control system with regards to cryptocurrencies that enhance the likelihood of achieving the entity's objectives relating to operations, reporting and compliance | Reporting Objectives         | FX risk of Bitcoin                                                           | 3.14                        | 3.66                | <ul style="list-style-type: none"> <li>- converting incoming Bitcoin funds instantly into fiat currencies</li> <li>- hedging of Bitcoin positions through certificates (in case of an investment)</li> <li>- appropriate documentation (amounts, FX rate, date and transaction details)</li> <li>- obsevation &amp; documentation of FX rate</li> </ul>                                                                                                                                                                                                              | pr/di<br>a/m<br>pr/di<br>a/m<br>pr/di<br>m                                          |
|                                                                                                                                                                                                              |                              | Financial reporting risks                                                    | 2.03                        | 2.96                | <ul style="list-style-type: none"> <li>- hiring a professional (external) advisor</li> <li>- appropriate documentation (amounts, FX rate, date and transaction details)</li> <li>- establishing guidelines for the national report of Bitcoin</li> <li>- establishing guidelines related to the monitoring of liquidity in Bitcoin markets</li> <li>- converting incoming Bitcoin funds instantly into fiat currencies</li> <li>- manual review of transactions/ financial reporting</li> </ul>                                                                      | pr/di<br>m<br>pr/di<br>a/m<br>pr/di<br>m<br>pr/de/cr<br>m                           |
|                                                                                                                                                                                                              |                              | Tax risks                                                                    | 2.18                        | 2.70                | <ul style="list-style-type: none"> <li>- appropriate documentation (amounts, FX rate, date and transaction details)</li> <li>- establishing guidelines for the recognition and valuation of Bitcoin transactions</li> <li>- hiring a professional (external) advisor</li> <li>- manual review of transactions/ paid taxes</li> </ul>                                                                                                                                                                                                                                 | pr/di<br>de<br>pr/di<br>m<br>pr/de/cr<br>m                                          |
|                                                                                                                                                                                                              |                              | KYC/AML Risks                                                                | 2.28                        | 2.95                | <ul style="list-style-type: none"> <li>- additional KYC check/ identification of customer as part of the ordering process</li> <li>- reporting of suspicious customers</li> <li>- internal assignment/ mapping of public addresses to (real) customers</li> <li>- establishing an internal IT forensic division</li> <li>- using data analysis tools to detect unusual transactions</li> <li>- monitoring of applicable national regulation and if necessary implementation of the relevant actions</li> </ul>                                                       | pr/de<br>a/m<br>di/de<br>a/m<br>di/de<br>a<br>pr/de/cr<br>a/m<br>pr/di<br>m         |





**Figure 23:** Cryptocurrency IC Framework; Source: Own illustration on the basis of the COSO 2013 IC Framework

incoming funds, such as hot or cold storage methods. Since cryptocurrencies are based on a new technology that has no other existing counterpart, a lack of knowledge would certainly aggravate every potential risk. Hence, it is imperative for a company to understand what a cryptocurrency is and how to deal with it in order to develop and further implement an appropriate risk management approach for its operational dealings.

2. **Identify & Assess:** In the second step, the potential risks that might emerge from the implementation of the new payment method have to be identified. For the risk identification process of cryptocurrencies, all potential threats that could influence the operational, reporting, and compliance objectives need to be analyzed and summarized (cf. the identified risks related to cryptocurrencies in Chapter 4.1). This process is continuous and not one-time because of ongoing changes in the company environment. However, when this process is "completed," the subsequent risk evaluation begins, which can be done by means of a risk assessment (cf. Chapter 4.2). By evaluating the risks previously identified according to their probability and severity, the risk assessment helps to determine the greatest risks.
3. **Develop Control Activities:** Thereafter, appropriate control activities have to be developed to mitigate the risks, particularly the most harmful ones, and these then need to be implemented. In this context, a risk control matrix (cf. Chapter 4.3.1) is a useful tool to help achieve this as it increases the transparency for all stakeholders and enhances their risk awareness, which in turn supports the process of developing control activities addressing the risks identified. These control

activities may vary in terms of their form (preventive, directive, detective, or corrective) and type (automatic or manual). In the case of cryptocurrencies, it is more reasonable to focus on automatic, preventive, and directive control activities due to the irreversibility of transactions and the resulting irrevocable losses.

4. **Inform & Communicate:** Nevertheless, it is not enough to simply implement an effective internal control system because the controls also have to be correctly applied and executed. Therefore, when the control activities are implemented, appropriate information and communication within the company (across several business units) is also required. This particularly represents a problem in the context of cryptocurrencies because the most important information about this complex technology-based payment method has to be selected. Moreover, the necessary information needs to be prepared and communicated in a way that any employee is able to comprehend the payment method, the risks associated, and the necessary measures needed to mitigate them.
5. **Monitoring Activities:** Monitoring plays a significant role in ensuring that the controls implemented are efficient and effective. This implies that internal measures like guidelines, demarcated responsibilities, etc., have to be monitored and, if necessary, adapted. Furthermore, the external environment (e.g., the emergence of new cryptocurrencies, changes in blockchain technology, new regulations, etc.) should also be continuously screened so as to identify newly emerging risk factors. If there are any changes or new risks recognized, the company needs to conduct a renewed risk assessment, which means that the circle will start anew.

The established Cryptocurrency IC Framework, such as the COSO 2013 IC Framework, serves as a tool for an organization's management to develop a reasonable internal control (system), particularly with regard using cryptocurrencies as a payment method. The framework, however, does need to be adjusted and adapted to the particular conditions of each company, such as its size, regulatory environment, or the share of Bitcoin payments. Also, one must always take into account that the framework increases the likelihood of the company achieving its objectives, but it gives no absolute assurance that the goals will be definitely accomplished or that the emerging risks of accepting Bitcoin funds will be completely eliminated.

## 5. Conclusion

### 5.1. Summary of results

The literature review shows that the implementation of cryptocurrencies has led to the emergence of new risks that might affect a company's operational effectiveness, reporting objectives, and compliance regulations. Due to this, its operational effectiveness could suffer from a lack of know-how, a segregation of duties concerning the administration of Bitcoin funds, IT-related risks, and issues like the irreversibility of transactions. The reporting objective might be affected by the various national reporting standards and the resultant misrepresentation of financial statements but also due to the considerable FX risk of bitcoins received. Further, the compliance objectives could be influenced by issues related to tax-compliance, AML, and KYC.

By confronting companies that accept Bitcoin payments with these risks, it could be determined that their risk awareness and the corresponding evaluation of the risks proposed is quite poor. Risks related to FX are currently experienced as the most significant, and cyber risks are also considered as serious threats. The administration of Bitcoin plays an important role here as both risks are highly interrelated, for instance, private key storage depends on the (internal IT-related) storage option chosen. In contrast, the irreversibility of transactions and emerging tax risks are less regarded as threats for the operational dealing of cryptocurrencies. Moreover, the analysis revealed that methods to deal with the risks of cryptocurrencies are not yet firmly established, and many corporations had a significant lack of know-how in this regard. This can be seen in the fact that less than half (46%) of the participating companies have implemented internal controls, and around 29% do not recognize a segregation of duties as a risk so as to ensure appropriate risk management concerning the operational dealing of cryptocurrencies.

Nevertheless, a majority of the companies (46%) using internal controls have implemented, presented effective control activities to mitigate the potential risks identified. On the basis of the results obtained through this study, the author has developed a feasible concept that can be further adapted and supplemented depending on each organization's size, risk complexity, and additional changes in the cryptocurrency environment. By considering the Cryptocurrency IC

Framework in combination with the integrated risk control matrix, an effective approach is proposed, aimed at ensuring the proper handling of cryptocurrencies and increasing the likelihood of an organization achieving its operational, reporting, and compliance objectives.

### 5.2. Discussion

The findings demonstrate that new, unequivocal risks emerge through the implementation of cryptocurrencies as a payment method, which can be due, on the one hand, to the properties of cryptocurrencies and on the other hand, to their underlying technology– the blockchain. However, the risks presented in this study may not be considered as a definitive list of all threats that might ever arise from using cryptocurrencies. In today's dynamic environment, a company must adapt to its environment, which requires the continuous monitoring and identification of newly emerging risks (Rüegg-Stürm and Grand, 2015, p. 78).

In general, the empirical analysis presented in this paper leads us to believe that companies do recognize these risks but only to a limited degree. Of course, only a small part of payments today are made using Bitcoin. This could be due to insufficient knowledge about the implementation of Bitcoin as a payment method but could also be due to its considerable increase in value during the last twelve months, which has made the cryptocurrency more attractive as a form of investment rather than a medium of exchange (cf. Appendix 2). However, as the share of crypto payments increases, its risk potential and severity evaluation will also become more important, because Bitcoin payments gain importance as their share in a company's total revenue increases. A logical implication of this is that the risk management involved in its operational dealing also becomes more important.

With regard to the risk awareness and risk assessment conducted, the FX risk was considered as the most important threat. One must mention that this risk might be completely eliminated by means of a payment processor and the immediate exchange into the relevant national currency. At first glance, the involvement of a merchant service provider simplifies the acceptance of Bitcoin payments and, moreover, several risks are already covered, such as the FX risk, the risk of payments defaulting, and, sometimes, KYC-related issues as well. Nevertheless, it is difficult to strike a balance between the advantages and disadvantages of a third-party provider because of the issue of third-party reliability and the processing fee per transaction. In this context, Massachusetts Institute of Technology's Professor Michael A. Cusumano (2014, p. 24) states that „payment processors and wallet firms are critical intermediaries and have also raised a lot of venture capital."

This statement indirectly introduces the second most significant risk, i.e., cyber risks, and the interrelated risk of the storage of bitcoins. Appropriate IT infrastructure and corresponding cyber security standards are critical for companies nowadays, especially in consideration of the storage of confidential data. As previously mentioned by Grant and Hogan (2015, p. 32), bitcoins are digital files "with a cash value

in the hands of the individual in possession of the file" or the corresponding private key. However, it is still difficult to limit the storage to the cold storage option because of the potential threat of the files being stolen by hackers. Cold storage also has drawbacks due to its physical nature. The recommendation of Sharma (2017, 14 November) and Antonopoulos (2015, p. 237), storing larger amounts of cryptocurrency assets on cold wallets and using hot wallets for day-to-day operations, seems reasonable. In this way, operational effectiveness could be improved and risks would be better diversified.

Independent of whether hot or cold storage method is chosen, it is important that the administration of Bitcoin payments is properly addressed. This can be done by limiting the access rights (to the private key or a hot wallet's password) to a small group of persons by assigning the responsibilities to those with the appropriate competence. Our findings indicate that, on average, two employees have the right to access this information, which seems appropriate, but one has to take into consideration that this analysis is mainly based on micro- and small-businesses and the results may differ in case of larger companies. Yet, there is no doubt that the field of competency should be assigned to someone within the senior management, preferably to someone who also has a high level of IT knowledge, like a CDO or CTO.

Apart from the fact that the risks are evaluated as somewhat low, the results of this study also indicate that there are currently a few companies with a significant lack of knowledge about cryptocurrencies. This became apparent in the respondents' answers like "I do not know where the company's bitcoins are stored" and "Everybody, or an intern, has the responsibility and access rights to manage the company's wallet." Additionally, the superficial risk evaluation considering the non-usage of a payment processor raises further issues. The answers show that the FX risk is considered as a threat by approx. 46% of the participants, so why do they not implement a payment processor that might mitigate this risk?

The last part of the analysis shows that the risks involved in dealing with cryptocurrencies are generally not yet firmly established because only around 46% of the companies claimed to have implemented internal controls in order to ensure their proper dealing. It is even more alarming that approx. 10% of the participants were not able to assess the compliance risks related to Bitcoin payments, and about 20% could not evaluate if the development of internal controls is possible in this regard. These findings support the assumption that there is an important lack of know-how within organizations.

### 5.3. Limitations & Suggestions for Future Research

Notwithstanding the results presented in the previous chapters, there are some limitations to this study.

To begin with, how cryptocurrencies and their usage as a payment method will evolve and develop cannot be anticipated. This is because today's environment is highly dynamic and constantly changing, so that it is almost impossible to determine concrete rules and regulations to control a disruptive

technology like the blockchain. Therefore, the results and the implied recommendations of this study are limited to the current status quo.

Second, the representative nature of this study is limited due to the selected sample, which mainly consists of micro- and small-businesses and thus does not represent the entire population (Brace, 2013, p. 7). This means that the results may differ for larger companies that have more complex structures that have a considerable impact on the assigned responsibilities and the related segregation of duties concerning the administration of Bitcoin funds within a company.

Furthermore, since Bitcoin payments currently only account for a small share of payments, it is important to note that the results of a risk awareness and evaluation may differ in the future. When the share of Bitcoin payments increases, it would also be required to conduct a more diligent reflection of the involvement of payment processors and the related costs and inherent exchange provider risks. The storage of Bitcoins will especially play an increasingly important role in this context. In addition, larger corporations located across several countries will be faced with the question of how to handle intercompany Bitcoin transfers because of differing national regulations and bans.

The scope of this study leaves open several options for future research. For instance, given the criticism of Bitcoin as a medium of exchange, future research could analyze cryptocurrencies that are a more appropriate fit as a medium of exchange, such as the zero-knowledge cryptography-based "Zcash" or the cryptocurrency "Ripple." The results of such studies could be used to adapt, specify, and supplement the Cryptocurrency IC Framework and the corresponding risk control matrix proposed.

Future studies could also adopt the opposite point of view by determining the general properties that any cryptocurrency should have in order to serve as an appropriate medium of exchange and, in particular, which risks should be addressed in the early stages of its development so as to prevent potential threats like how to make it a more secure medium of exchange.

## References

- Antonopoulos, A. M. *Mastering Bitcoin: unlocking digital cryptocurrencies*. "O'Reilly Media, Inc.", 2015.
- Arwinge, O. *Internal control: A study of concept and themes*. Springer Science & Business Media, 2012.
- Azzara, C. V. *Questionnaire design for business research: Beyond linear thinking-an interactive approach*. Tate Publishing, 2010.
- Bank for International Settlement [BIS]. Digital currencies [report], 2015. URL <https://www.bis.org/cpmi/publ/d137.htm>.
- Berschens, R., J., H., and Holtermann, F. Kampf um den Bitcoin, 2017, 19. December. URL <http://www.handelsblatt.com/my/finanzen/maerkte/devisen-rohstoffe/regulierung-der-digitalwaehrung-kampf-um-den-bitcoin/20742374.html>.
- Bhaskar, N. D. and Chuen, D. L. K. Bitcoin mining technology. In *Handbook of Digital Currency*, pages 45–65. Elsevier, 2015.
- Bitcoin Foundation. Developers, 2017. URL <https://bitcoinfoundation.org/developers/>.
- Blasius, J. and Baur, N. *Handbuch Methoden der empirischen Sozialforschung*. Springer, 2014.
- Blockchain. Transactions per Day, 2018, 20. January. URL <https://blockchain.info>.
- Böhme, R., Christin, N., Edelman, B., and Moore, T. Bitcoin: Economics, technology, and governance. *Journal of Economic Perspectives*, 29(2):213–238, 2015.
- Brace, I. Questionnaire design. how to plan, structure and write survey material for effective market research. London: Kogan Page Ltd., 2013.
- BTC Echo. Bitcoin-Akzeptanzstellen, 2017. URL <https://www.btc-echo.de/bitcoin-akzeptanzstellen>.
- Buber, R. and Holzmüller, H. H. Qualitative Marktforschung: Konzepte-Methoden-Analysen, 2. Aufl., Wiesbaden, 2009.
- Bungartz, O. *Interne Kontrollsysteme (IKS): Basiswissen für den Aufsichtsrat*. Erich Schmidt Verlag GmbH & Company, 2017.
- Casari, E. Internal Control and Fraud Prevention (Master's Thesis, University of St. Gallen), 2017. URL [http://www1.unisg.ch/org/biblio/edoc.nsf/SysLkpByIdentifier/10608826102/\\$FILE/10608826102.pdf](http://www1.unisg.ch/org/biblio/edoc.nsf/SysLkpByIdentifier/10608826102/$FILE/10608826102.pdf).
- Central Intelligence Agency. The World Factbook - Country Comparison: Stock of Narrow Money, 2018. URL <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2214rank.html#rs>.
- Choo, K.-K. R. Cryptocurrency and virtual currency: Corruption and money laundering/terrorism financing risks? In *Handbook of digital currency*, pages 283–307. Elsevier, 2015.
- Cohen, E. The Bitcoin Economy. *Business NH Magazine*, 31(6):18–19, 2014.
- Coinbase. Accept Bitcoin Payments, 2018, 12. February. URL <https://www.coinbase.com/merchants>.
- CoinDesk. Bitcoin Price Index, 2018, 20. January. URL Retrieved from <http://www.coindesk.com/price/>.
- Coinmap. Coinmap.org, 2017, 21. September. URL <https://coinmap.org/#/world/48.09275716/3.73535156/5>.
- Coinmap. Coinmap.org, 2018, 5. January. URL <https://coinmap.org/#/world/48.09275716/3.73535156/5>.
- Coinmarketcap. Cryptocurrency Market Capitalizations, 2018, 20. January. URL <https://coinmarketcap.com>.
- Coinmarketcap. Cryptocurrency Market Capitalizations, 2018, 4. February. URL <https://coinmarketcap.com>.
- COSO, C. o. S. O. o. t. T. C. Enterprise Risk Management - Integrated Framework. Executive Summary, 2004. URL <https://www.coso.org/Documents/COSO-ERM-Executive-Summary.pdf>.
- COSO, C. o. S. O. o. t. T. C. Leveraging COSO across the three lines of defense, 2013. URL <https://www.coso.org/Documents/COSO-2015-3L0D.pdf>.
- COSO, C. o. S. O. o. t. T. C. Internal Control – Integrated Framework. Executive Summary, 2015. URL <https://www.coso.org/Documents/990025P-Executive-Summary-final-may20.pdf>.
- Cusumano, M. A. The bitcoin ecosystem. *Communications of the ACM*, 57(10):22–24, 2014.
- de Jong, I. The role of virtual currency schemes in the modernisation of retail payments. *Journal of Payments Strategy & Systems*, 8(4):415–425, 2015.
- Deloitte. Vorstellung der Blockchain-Technologie, 2016. URL <https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Innovation/Vorstellung%20der%20Blockchain-Technologie.pdf>.
- Di Pierro, M. What is the blockchain? *Computing in Science & Engineering*, 19(5):92–95, 2017.
- Dombrowski, E. Analysis of bitcoin as a viable currency, 2014. URL [http://www1.unisg.ch/org/biblio/edoc.nsf/SysLkpByIdentifier/10603116101/\\$FILE/10603116101.pdf](http://www1.unisg.ch/org/biblio/edoc.nsf/SysLkpByIdentifier/10603116101/$FILE/10603116101.pdf).
- ECB, E. Virtual currency schemes, 2012. URL <https://www.ecb.europa.eu/pub/pdf/other/virtualcurrencyschemes201210en.pdf>.
- Elo, S. and Kyngäs, H. The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1):107–115, 2008.
- Flick, U. Triangulation. Eine Einführung. 3., aktualisierte Auflage, 2011.
- Franco, P. *Understanding Bitcoin: Cryptography, engineering and economics*. John Wiley & Sons, 2014.
- Gautham. Japan Officially Recognises Bitcoin as Currency Starting April 2017. NewsBTC, 2017. URL <http://www.newsbtc.com/2017/04/02/japan-officially-recognises-bitcoin-currency-starting-april-2017/>.
- Gisler, M. *Virtuelle Währungen-Eine ökonomische und finanzmarktrechtliche Einordnung am Beispiel Bitcoin*. PhD thesis, Universität St. Gallen, 2015. URL [http://www1.unisg.ch/org/biblio/edoc.nsf/SysLkpByIdentifier/08604548102/\\$FILE/08604548102.pdf](http://www1.unisg.ch/org/biblio/edoc.nsf/SysLkpByIdentifier/08604548102/$FILE/08604548102.pdf).
- Goldman Sachs. Blockchain – Putting Theory into Practice [Equity Research Report], 2016. URL [https://www.reddit.com/r/BlockChain/comments/58xf3q/goldman\\_sachs\\_report\\_may\\_2016\\_blockchain\\_putting/](https://www.reddit.com/r/BlockChain/comments/58xf3q/goldman_sachs_report_may_2016_blockchain_putting/).
- Graham, L. *Internal control audit and compliance: documentation and testing under the new coso framework*. John Wiley & Sons, 2015.
- Grant, G. and Hogan, R. Bitcoin: Risks and controls. *Journal of Corporate Accounting & Finance*, 26(5):29–35, 2015.
- Greene, C. and Shy, O. E-cash and virtual currency as alternative payment methods. *Journal of Payments Strategy & Systems*, 8(3):274–288, 2014.
- Heer, C. Bilanzierung von Bitcoins, 2014, 10. December. URL <https://www.wadsack.ch/news/datum/bilanzierung-von-bitcoins/>.
- Hern, A. A history of bitcoin hacks. The Guardian, 2014, 18. March. URL <https://www.theguardian.com/technology/2014/mar/18/history-of-bitcoin-hacks-alternative-currency>.
- Hoelscher, J. L. Digital currency risks: internal auditors need to advise management on the myriad risks posed by new forms of payments. *Internal Auditor*, 71(4):24–26, 2014.
- Hunziker, S. *Erfolg der Internal Control-Eine empirische Analyse aus Sicht des Managements*. PhD thesis, Dissertation, Universität St. Gallen, St. Gallen, 2015.
- IMF, I. Virtual Currencies and Beyond: Initial Considerations, 2016. URL <https://www.imf.org/en/Publications/WP/Issues/2017/03/27/The-Macroeconomics-of-De-Cashing-44768>.
- IMF, I. The Macroeconomics of De-cashing, 2017. URL <https://www.imf.org/external/pubs/ft/sdn/2016/sdn1603.pdf>.
- Kanton, Z. Kryptowährungen (Bitcoin, Ethereum, Tokens usw.) [Merkblatt Steuern für Privatpersonen], 2017. URL <https://www.zg.ch/behoerden/finanzdirektion/steuerverwaltung/kryptowaehrungen>.
- Kollewe, J. Bitcoin: UK and EU plan crackdown amid crime and tax evasion fears. The Guardian, 2017, 4. December. URL <https://www.theguardian.com/technology/2017/dec/04/bitcoin-uk-eu-plan-cryptocurrency-price-traders-anonymity>.
- Lam, J. *Enterprise risk management: from incentives to controls*. John Wiley & Sons, 2014.
- Lindner, T. and Meyer, S. D. Besteuerung und Bilanzierung von Bitcoin. *Rechnungswesen & Controlling*, Heft 4/2017, 28-29, 2017.
- Lloyds. Emerging Risk Report 2015 – Technology: Bitcoin, 2015. URL <https://www.lloyds.com/news-and-risk-insight/risk-reports/library/technology/bitcoin>.
- Morettini, P. Does Blockchain Fit in Your Technology Stack?, n.D. URL <https://www.pjmconsult.com/index.php/2018/01/blockchain-technology-stack-software.html>.
- Nakamoto, S. A Peer-to-Peer Electronic Cash System, 2008. URL <https://bitcoin.org/bitcoin.pdf>.
- Nguyen, J. The Risks of Segregated Witness: Problems under Electronic Contract and Evidence Laws. *Computer & Internet Lawyer*, 34(11):1–8, 2017.
- Nian, L. P. and Chuen, D. L. K. Introduction to bitcoin. In *Handbook of Digital*



- Currency, pages 5–30. Elsevier, 2015.
- Nofer, M., Gomber, P., Hinz, O., and Schiereck, D. Blockchain. *Business & Information Systems Engineering*, 59(3):183–187, 2017.
- of Internal Auditors (IIA), I. The three lines of defense in effective risk management and control. *Position paper*, 2013.
- Papadopoulos, G. Blockchain and Digital Payments: An Institutional Analysis of Cryptocurrencies. In *Handbook of Digital Currency*, pages 153–172. Elsevier, 2015.
- Pfaff, D. and Ruud, T. F. Schweizer Leitfaden zum internen Kontrollsystem (IKS). Verlag: Orell Fuessli. Switzerland, 2013.
- Piller, F. Virtuelle Währungen - Reale Rechtsprobleme? Eine zivil-, zwangsvollstreckungs- und steuerrechtliche Untersuchung aus der Sicht der Bitcoin-Nutzer, 2017. URL [http://www1.unisg.ch/org/biblio/edoc.nsf/SysLkpByIdentifier/12608618102/\\$FILE/12608618102.pdf](http://www1.unisg.ch/org/biblio/edoc.nsf/SysLkpByIdentifier/12608618102/$FILE/12608618102.pdf).
- Rouse, M. Database (db), 2017. URL <http://searchsqlserver.techtarget.com/definition/database>.
- Rüegg-Stürm, J. and Grand, S. *Das St. Galler Management-Modell*. Bern: Haupt, 2015.
- Ruud, F. and Jenal, L. Licht im Internal Control-Dschungel. *Der Schweizer Treuhänder*, 79(6-7):455–460, 2005.
- Ruud, F., Isufi, S., and Friebe, P. Pflicht zur Prüfung der Existenz des Internen Kontrollsystems: Bestandesaufnahme zur Steuerung und Kontrolle mittelgrosser Unternehmen in der Schweiz. *Der Schweizer Treuhänder*, 82(11):938–942, 2008.
- Schmidt, T. Kryptowährungen und Steuern: Was jetzt wichtig wird. BTC-Echo, 2018, 18. January. URL <https://www.btc-echo.de/kryptowae-hrungen-und-steuern-was-jetzt-wichtig-wird/>.
- Sharma, T. K. Hot wallet vs cold wallet. The Blockchain Council, 2017, 14 November. URL <https://www.blockchain-council.org/cryptocurrency/hot-wallet-vs-cold-wallet/>.
- Sixt, E. Bitcoins und andere dezentrale Transaktionssysteme. *Blockchains als Basis einer Kryptoökonomie*, Wiesbaden: Springer, 2017.
- Sue, V. M. and Ritter, L. A. Conducting online surveys. Thousand Oaks, CA: SAGE Publications Ltd. 2007.
- Swan, M. Blockchain. blueprint for a new economy: Beijing: O'Reilly, 2015.
- Tindell, K. Geeks Love The Bitcoin Phenomenon Like They Loved The Internet In 1995. Business Insider, 2013, 5. March. URL <http://www.businessinsider.com/how-bitcoins-are-mined-and-used-2013-4?IR=T>.
- Underwood, S. Blockchain beyond bitcoin. *Communications of the ACM*, 59(11):15–17, 2016.
- United States Treasury Inspector General for Tax Administration (TIGTA). As the Use of Virtual Currencies in Taxable Transactions Becomes More Common Additional Actions Are Needed to Ensure Taxpayer Compliance [report], 2016. URL <https://www.treasury.gov/tigta/auditreports/2016reports/201630083fr.pdf>.
- Vigna, P. and Casey, M. The Age of Cryptocurrency: How Bitcoin and the Blockchain are Challenging the Global Economic Order: New York: Picador, 2016.
- Westerhausen, H.-U. Das COSO-Modell: bisher nur eine Randerscheinung in Deutschland? Ein 13 Jahre altes IKS-Modell und seine Verbreitung in Deutschland. *Zeitschrift Interne Revision*, (3):98–103, 2005.
- Worldcoinindex. Overview, 2018, 9. February. URL <https://www.worldcoinindex.com/trending/overview>.





# Does Subordinated Debt Discipline Banks? Empirical Evidence of Market Discipline in Europe

Daniel Schürk

*Johann Wolfgang Goethe Universität*

## Abstract

This thesis provides a differentiated answer to the question whether subordinated debt disciplines bank's risk-taking behavior. I investigate the conditions and applicability of market discipline through subordinated debt instruments by critically reviewing the state of research. Relating to the regulatory context, I discuss proposals and various empirical studies and find that subordinated debt is an adequate measure to discipline banks under certain conditions.

My own empirical analysis contributes to evidence provided by prior studies and updates them for the European case. I conclude that subordinated debt investors perceive differences in risk between banks and across time and are sensitive to credit ratings and accounting variables at generally higher spread levels compared to senior bonds. Results include that spread is positively sensitive (increases with respect to one standard error) to equity to capital (225 BPS), provision for loan losses (200 to 225 BPS), non-performing loans to equity (400 to 715 BPS) and interest coverage ratio (60 BPS). Spread is negatively sensitive (decreases with respect to one standard error) to ROA (120 BPS) and loan loss reserves (360 to 620 BPS).

**Keywords:** debt market discipline; bond spreads; subordinated debt; bail-in; bail-out; BRRD; Basel II; Basel III; market monitoring; market influence.

## 1. Introduction

During the Great Financial Crisis ("GFC"), tax-payer's money was repeatedly used to bail out banks considered 'too big to fail' and as such seen as a relevant risk to the financial and economic system (Barth and Wahlborg, 2016). It was widely criticized that banks did not bear the costs of their own risks anymore and that risks were not controlled properly (Hakenes and Schnabel, 2010). Underestimating the fragility, interconnectedness and complexity of global banks led regulators to weaken supervisory efforts and elevated risk-taking behavior of banks in the years preceding the crisis. In the aftermath of the crisis, however, regulation in the banking sector and especially in Europe tightened significantly to mitigate excessive risk-taking and to avoid further costly bail-outs (Dewatripont, 2014). Exclusively relying on direct regulation measures, such as capital and disclosure requirements, might not be sufficient to cope with banks' risks, as noted in the Squam Lake Report (French et al., 2010).

Lately, regulation is increasingly putting emphasis on the potential of self-controlling, market-based mechanisms, termed 'market discipline' ("MD"). The rationale is that by monitoring bank's behavior and condition, market par-

ticipants are ultimately influencing risk-taking reciprocally when managers take into account capital market's evaluation (Flannery, 2001). This concept could supplement traditional supervisory measures (Hoang et al., 2014). Research on market discipline of debt capital has expanded with increased interest from supervising authorities. Following a surge in publications in the years around the turn of the millennium, one can notice that more research has been provided since 2009 paralleled by ongoing negotiations on Basel III. With the new European framework on Bank Recovery and Resolution ("BRRD"), regulators are committed to strengthen market discipline further (Lintner, 2017), by implementing a bail-in tool.

Central to the concept of market discipline is the asset class of subordinated debt ("SD")<sup>1</sup>. These high-yield instruments rank just after equity when it comes to absorbing losses in capital restructuring ('bail-in') and are thus exposed to higher risk of default. When reflected in the risk premium on markets (as in spreads), this should provide incentive to investors to monitor default risks (Chen and Hasan, 2011).

<sup>1</sup>More specifically, I will only refer to subordinated bonds in this thesis.

In theory, this quality makes them an ideal tool for exerting MD, serving as an indicator for investor's sensitivity to risk, when perceiving the bank's risk (Bliss and Flannery, 2002). Higher yields are demanded by investors and account for riskier projects and assets of the bank. I want to test this rationale empirically to answer the question whether SD is an adequate instrument to discipline banks in Europe. I will discuss conditions and implications constituting market discipline and the channels through which SD might discipline banks. Further analyzing the monitoring and influencing phases, I will highlight limitations to the implementation of MD and present proposals which respond to these issues. In the main body of my thesis, the empirical work, I will present my own study on the relevance and sensitivity of SD in Europe. My analysis will contribute to the current discussion since evidence from earlier studies could be outdated due to regulatory adjustments and higher commitment to banking resolution in Europe.

This thesis seeks to answer the following main research questions:

1. What is the context and rationale of subordinated debt when establishing market discipline in the banking sector?
2. Under which conditions can subordinated debt effectively work as an adequate measure to monitor and influence a bank's risk-taking behavior?
3. What are the features and conflicts to the implementation of subordinated debt presented in current proposals?
4. What are the characteristics of the current market for subordinated debt in Europe?
5. Is there empirical evidence of sensitivity of subordinated bond spreads towards the risk profile of a bank, i.e. can investors monitor the bank?

Narrowing the objective of my study, the empirical part of the thesis will clearly focus on the monitoring phase of SD, that is the sensitivity of spreads to risk measures, and will not aim at providing evidence for the influencing phase. Moreover, I will not discuss the general comparison on the rationale of bail-ins versus bail-outs, the role of deposit insurance, legal details of SD or MD or models on moral hazard behavior in banks.

The remainder of this paper is structured as follows: the following section (section 2) introduces the regulatory context and core concepts and briefly reviews the literature on implications, drawbacks and conditions for effective market discipline before presenting proposals, addressing research questions 1 to 3. Section 3 discusses research questions 4 and 5 and provides a review of empirical evidence on spread sensitivity of SD, methodological considerations and my own empirical results for the European case. Section 4 concludes the thesis with a summary, open research questions and an outlook.

## 2. Analysis of effective market discipline by subordinated debt

This section builds the foundation for understanding the debate on SD as a disciplining instrument. A critical analysis will be undertaken on recent and earlier research papers to evaluate the viability of SD as a measure for disciplining banks.

### 2.1. Regulatory context

The regulatory framework of banking supervision is originally designed to protect the economy from systemic risk. However, when governments also act as a lender of last resort and provide a deposit insurance system, this could entail incentives for banks to increase leverage and risk-taking respectively (Flannery, 2001). According to his findings, government guarantees and deposit insurance are mispriced, meaning that banks and their owners do not bear their own risk. Additionally, there could exist higher social cost from public oversight compared to private governance systems in the banking sector. These aspects, combined with the repeated inability of supervisors to mitigate banking crises, as shown in Lall (2012), call for alternative supervisory mechanisms.

After many financial crises (more specific: banking crises) regulators are thus adjusting their framework, imposing stricter regulation on capital adequacy and disclosure (Barth et al., 2004). On a supranational level, banking regulation was established by the Basel programs with the objective of improving transparency and reliability of financial information of banks. Most of these measures are focused on the prevention of public bail-outs, prevalent in the GFC, which can easily spark a chain reaction as a spillover effect on a public debt crisis. Conversely, if regulation is designed too tight in terms of capital requirements, this could contract the vital function of banks creating liquidity (Van den Heuvel, 2008).

To overcome this dilemma of direct regulatory requirements, the third pillar of Basel II endeavors to implement alternative regulatory measures, making use of market instruments as a self-controlling mechanism to let the market discipline banks partially. Optimally, this releases governments from the inherent paradigm of giving guarantees to bail out banks (Grossman and Woll, 2014) and returns the accountability for losses back to markets, i.e. investors who are bailed-in depending on the payment rank of their capital. The newly implemented Banking Union in the EU and the recovery and resolution mechanism for banks (BRRD), turned into effective law in the national states by 2016, supports the internalization of liability. Resolution mechanisms in BRRD (Lintner, 2017) prescribe a clear ranking of bail-in procedure with equity holders absorbing losses prior to unsecured junior debt (so-called waterfall principle). Current European regulation under the name Minimum Requirement for Own Funds and Eligible Liabilities ("MREL") specifies the criteria for subordination and bail-in order following the international rules for Total Loss Absorbing Capacity ("TLAC")

introduced by the Financial Stability Board. Beyond the commitment to a bail-in tool, the new regulation wants to evade TBTF guarantees. In June 2017, for the first time since the establishment of BRRD, subordinated debt holders and shareholders of struggling Banco Popular, the sixth-largest Spanish bank, were bailed in and absorbed full losses without harming tax-payers (Buck, 2017). This might serve as a blueprint to further resolutions, possibly in sight regarding the condition of troubled Italian banks (Martino, 2017).

While the bail-in order is not a self-purpose or a mechanism for quick resolution only, it is also intended to have repercussions on the awareness of risk to mitigate the impact of future banking crises (Dewatripont, 2014) eventually. 'Bail-inable' capital is thus elevating market discipline.

## 2.2. The case for market discipline

Market discipline can be defined in its broadest sense as a mechanism by which bank's risk-taking behavior is monitored and influenced by market forces (Hamalainen et al., 2009). Commentators are arguing to which extent MD might supplement and replace regulators and if it does so at all (Bliss et al., 2001). Whereas the notion of liability and fundamentals of MD are already common for non-financial corporations, implementation in a heavily regulated financial environment seems to be more complex. In a speech delivered at the New York University Law School in April 2007 (see quote on next page), the then-chairman of the Fed, Ben Bernanke, was already aware of the need for MD. It should be noted that this speech was held just few months before the beginning of the crisis – not knowing what was about to happen in the succeeding months.

"The lesson of history appears to be that neither market discipline nor regulatory oversight alone is completely adequate for keeping the banking system safe and sound. However, regulators have increasingly come to appreciate the value of a hybrid system that supplements direct regulation with a substantial amount of market discipline. Fortunately, regulators have a variety of ways to restore and strengthen market discipline for banks, notwithstanding the existence of the federal safety net." (Bernanke, 2007)

However, the recent incorporation of MD in the banking sector (Bernanke's hybrid system) is not a new phenomenon. Private governance frameworks and incentive-based regulatory design emerged in the US (Bliss and Flannery, 2002) and are generally proposed by opponents of too strict government controls, since they give more weight to market-based mechanisms. Discussions were initially sparked in a similar context to the present post-GFC period, when several US banks defaulted during the savings and loan crisis in the 80s of the last century as government guarantees led them to higher risk-taking (Board of Governors of the Fed, 1999).

Besides concepts of MD focusing on equity and deposits, the "Squam Lake Report" repeatedly stressed the relevance

of debt market discipline (French et al., 2010). This form of MD can be broken down into two phases: the monitoring phase followed by the influencing phase. Bliss and Flannery (2002), among other authors, differentiate between the ability to understand, interpret and price information on changes to the banks' decisions (monitoring) and the ability to ultimately impact the decision maker's behavior via security prices (influence).

For the influencing phase, the common literature distinguishes further between direct and indirect disciplinary effects (Hamalainen et al., 2009). Direct influence relates to higher costs of borrowing for new issuances on primary markets through a higher risk premium. Investors are anticipating the risk of default banks and can thus influence bank managers behavior ex-post. Indirect discipline refers to the regulator, who is interpreting and incorporating market signals from secondary markets into its supervisory actions.

## 2.3. Why subordinated debt?

Until now, I only discussed MD in its theoretical rationale. Both direct and indirect MD, however, require an adequate instrument on markets, where expectations are traded and priced. While MD might be also imposed by shareholders or relevant depositors (Demirgüç-Kunt and Huizinga, 2004), debt markets are particularly suitable for implementing MD due to higher risk sensitivity of debt holders and generally a higher frequency of debt issuances (Chen and Hasan, 2011). Constant payments on the face value encompass a downside risk without the benefit of large upside gains, common for equity holders.

SD is typically unsecured capital and uninsured by any governmental safety net. It can also come with convertible features; however, I will regard CoCos only as a sub-category to SD. Basel III regulation accounts SD to Additional Tier 1 if it is subordinated to the general debt and if it has no maturity (perpetual bonds) and to Tier 2 capital if maturity is longer than 5 years<sup>2</sup>. It thereby ranks lower in the liquidation scheme than normal debt in case of default and acts as an additional capital buffer similar to equity capital, supplementing Tier 1 capital and increasing the robustness of deposit insurance besides strengthening the position of senior debt holders (Caldwell et al., 2005). What makes SD particularly interesting to banks is its simple and quick issuance compared to stocks. Investors, on the other hand, are attracted by higher yields.

Being exposed to a higher risk profile, reflected in the risk premium on markets (as in spreads), should provide more incentive to investors to monitor default risks (Chen and Hasan, 2011). In turn, this price sensitivity of SD possibly carries valuable information about default risks and might better reflect the financial situation of a bank compared to senior debt. SD is thus a better instrument for MD compared to depositors, who can silently withdraw their funds and do not provide an immediate market signal (Calomiris, 1999).

<sup>2</sup>See BCBS publication 189 for Basel III rules on capital treatment, available at <http://www.bis.org/publ/bcbs189.pdf>

## 2.4. Conditions for effective market discipline

Reflecting Bernanke's speech quoted above from today's standpoint, the obvious question is why MD failed during the crisis. Since bond spreads seemed to react only when it was too late, we would have to ask which conditions were not met at that time prior to the crisis. In this section, I will discuss conditions and limitations to MD. For further analysis, I will refer to the comprehensive framework graphed in Figure 1, which is an adapted representation from Hamalainen et al. (2009).

### 2.4.1. Market monitoring

The market monitoring phase (or recognition phase in Hamalainen et al., 2009) can be seen as the necessary pre-phase for the succeeding phase of market influence (see section 2.4.2). Market participants should be able to understand changes in a bank's condition (sub-phase 1 in the above framework), consider themselves at risk (sub-phase 2) and price this perception accordingly (sub-phase 3). Several studies show that bond prices are indeed related to the condition of banks or bank holding companies, especially prior to bank failure (Jagtiani and Lemieux, 2001)<sup>3</sup>. There are, however, several conditions which must be fulfilled to allow investors to perceive and price risk accordingly.

First, to satisfy condition C1, the 'informed investor' hypothesis must hold (Hancock and Birchler, 2004). Information gathering must not be linked to high costs and there must be constant information provision towards investors. This requires a regulatory framework with high transparency. The strengthened reporting and disclosure rules together with credit ratings should set the baseline for informed markets. Nevertheless, the reliability of the signal of subordinated bond yields might be limited due to infrequent issuing behavior (Evanoff et al., 2011). Admati and Hellwig (2013) even call debt market discipline a myth due to high informational costs incurred by the creditors. Investors could well be 'free-riding' on information provided by other investors which, in effect, lowers the incentive for monitoring. Moreover, the efficacy of monitoring also depends on the type of investor and their ability to detect risks. While institutional investors or peer banks might actively monitor banks and financial status, private investors incur much higher information costs for monitoring internal risks. For example, in the Italian SD market, half of the amount issued is allocated to individual investors (see Martino, 2017). This undermines the 'informed investor' assumption further.

Regarding the second condition (C2), the 'too-big-to-fail' argument presented in Mishkin et al. (2006) undermines the relevance of MD since investors anticipate being bailed out by the safety-net of the government. For effective MD, however, explicit and implicit guarantees for public bail-out must be irrelevant to investors so that they bear the full risk of losses on their investments and thus internalize the bank's risk. In

the current European context, this condition is increasingly enforced through the BRRD framework.

To further ensure efficiency (condition 3), markets must be sufficiently deep in terms of liquidity. Although this might be fulfilled for large banks and issuances, Chen and Hasan (2011) note that SD might not be the right instrument of choice for monitoring small banks, since low liquidity on these markets (especially on secondary markets) comes with increased noise and volatility. The lower trading volume on secondary markets leads to a liquidity premium one would have to adjust for (Evanoff et al., 2011).

Another pitfall to monitoring goes beyond the three conditions. The so-called 'double endogeneity problem' presented in Birchler and Facchinetti (2007) questions the ability of markets to price risk when participants assume that banks are disciplined by their own monitoring efforts. They anticipate their own disciplining effect into their pricing process and do not exert discipline consequently.

### 2.4.2. Market influence

The ability to monitor risk solely is not sufficient for holistic MD. The signal provided must also be put into actionable controls, either by creditors (direct MD) or by the supervisor (indirect MD).

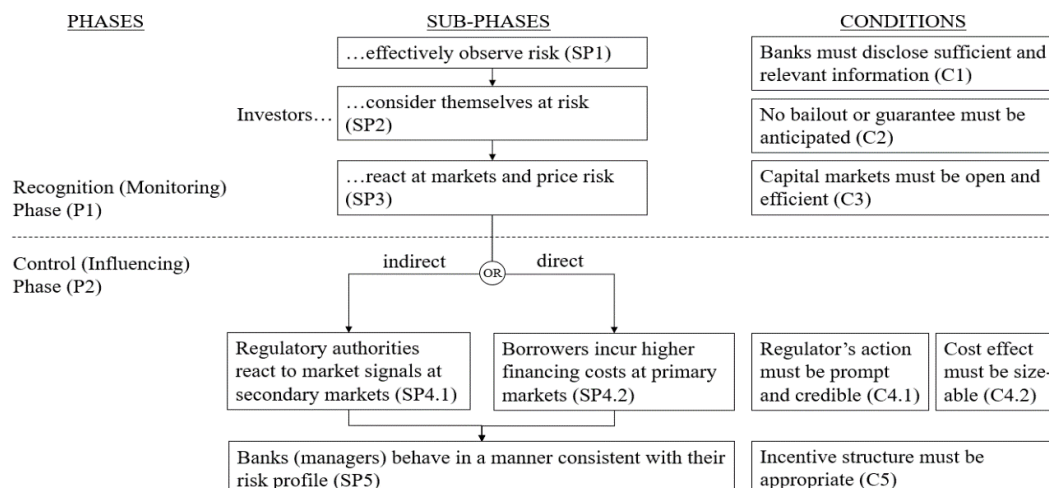
Direct market discipline ("DMD") is exerted directly by markets when demanding a higher risk premium. Hwang and Min (2013) distinguish DMD further between weak and strong MD. The former is referring to higher financing costs incurred when investors are pricing the risk accordingly; the latter is reflecting the inability to raise further capital if the probability of default surpasses a critical line. Strong MD will probably only be observed in times of approaching distress. Besides to the pure financing view, one might also expect an ex-ante effect of influence imposed by covenants as another channel of MD, as shown by Goyal (2005). Since the strength of covenants of different instruments is negatively related to the risk premium, there can surely be observed different spreads across instruments.

The Board of Governors of the Fed (1999) emphasizes that the share of SD must be relevant and sizeable (Condition 4.2) to impact total funding costs and influence managerial behavior. Bliss and Flannery (2002) provide research regarding this question on cost effects of spreads. They generally find that market prices alone are not strong enough to influence key decisions like the choice of leverage, dividends and the level of liabilities, probably due to the small market size of SD (breaching C4.2 in the framework).

There remains another pitfall to the use of SD regarding incentive structures (condition 5) which can be easily understood when considering SD as a capital source with similar characteristics as equity, particularly vulnerable to moral hazard problems. In the case of default, or rather close to that state, SD investors would as well support risky investments and be supportive for so-called 'gambling for resurrection' decisions (Bliss et al., 2001). In this case, the impact on spreads would be distorted.

<sup>3</sup>Refer to section 3.1 for an overview over other prior studies.





**Figure 1:** The phases of effective market discipline, adapted from Hamalainen et al. (2009)

SD spread signals attained from monitoring could also indirectly influence banks when anticipating regulatory action, which should be timely and credible (see condition C4.1). In contrast to DMD, indirect market discipline can only be exercised through publicly placed instruments as the pricing by markets must send an observable signal to the regulator. Although this market signal means a partial delegation of monitoring activities from the supervisor to markets, the supervisor will still step in to perform prompt corrective action or on-site examinations if distress is expected.

Current early-warning models and triggers for regulatory action primarily focus on book variables and the so-called CAMELS<sup>4</sup> rating to evaluate distress. Conversely, SD spreads could also be an adequate measure to incorporate market-based information on distress (Jagtiani and Lemieux, 2001). They could even be better predictors than traditional Tier 1 capital ratios (Evanoff and Wall, 2000). However, the noise of SD spreads (vide supra) makes them hard to interpret as a stable indicator; additional factors might naturally distort market data (see Chen and Hasan (2011) and Bliss et al. (2001)). One can also categorically question whether yields really comprise additional information to supervisors. Bliss et al. (2001) points out that investors only observe publicly available information also available to supervisors. He questions the signaling effect of bond spreads and advises to use combined measures, such as incorporating equity prices into risk assessment. Nevertheless, even an imprecise signal could complement supervisor's assessment and actions and could help them to react promptly to banks in distress (Flannery, 2001).

Summing up the previous sections, I refer to the conclusion of Flannery (2001), who supposes that SD spreads might inform about different risk levels (monitoring phase)

but do not necessarily discipline banks accordingly (influencing phase). In his view, supervisors are better at influencing firms and investors have a comparative advantage in monitoring banks via markets. To minimize the probability of failure in supervision, he calls for a combination of both channels.

## 2.5. Mandatory subordinated debt proposals

Due to high financing costs, banks are inherently reluctant to issue more SD voluntarily. Frequency of issuance and amount issued might rest at low levels in the future. Besides using market information from SD spreads only as evaluative information, authorities could turn to mandatory regulation to strengthen the role of MD, as desired in the speech by Bernanke quoted above. To enforce mandatory proposals, the regulator could require predefined levels of SD, for example calculated as a share of total debt, and act if banks are reluctant to issue sufficient SD (Chen and Hasan, 2011).

In the following, I will summarize only the most relevant proposals of mandatory programs and link them to the framework in Figure 1. For an extensive review of mandatory SD proposals, one can refer to a 'Staff Study' by Fed Board members (Board of Governors of the Fed, 1999). For an overview of proposals on features like maturity, size and further SD features, refer to Hamalainen (2004).

Early proposals emerged in the US with the Gramm-Leach-Bliley Act from 1999, which led the Fed to examine a mandatory program. However, the Board did not see extended need to take prompt action. They called for further research on smaller banks' barriers to issue and delayed a decision due to increasing numbers of voluntary issuances (Board of Governors of the Fed, 1999).

Calomiris (1999), as an outstanding advocate for the mandatory solution, puts forward a radical proposal. While a cap on spreads at issuance would serve to account for excessive risk-taking behavior, he also proposes a minimum frequency of issuance to provide markets with a constant stream

<sup>4</sup>Rating system to classify a bank's overall condition, evaluating (C)apital adequacy, (A)ssets, (M)anagement Capability, (E)arnings, (L)iquidity and (S)ensitivity.



of information (addressing condition C1). This would, however, require maturities to decrease far below current levels. If banks do not comply with the frequency of issuance or if bonds hit the spread cap, he proposes an obligation to shrink the asset side monthly to meet requirements.

Caldwell et al. (2005) picks up the debate on desired minimum maturity of SD instruments. He states that short-term SD does not anticipate long-term risks and would additionally come with higher costs of issuance given the higher frequency. The positive effects of liquidity come at a price. Both banks and supervisors might be willing to tend to longer maturities, seeking for permanence and stability. Long maturities, however, increase the risk to debt holders, tying them to the bank. The maturity requirements are therefore a pure trade-off between liquidity and information provision (monitoring phase) and the strength of the disciplining effect (influencing phase).

Caldwell et al. (2005) also points out that mandatory programs could prevent small banks from timing their issuance decision. By studying the SD market for large and small Canadian banks, he assesses an incentive for small banks to systematically place SD in 'quiet times' while avoiding issuance when spreads are higher. They might also issue SD before negative news on their risk-taking behavior reach the market and thus actively evade MD. In contrast to big banks, small banks seem to be more inclined to time their issuing behavior due to higher cost of funding. He suggests a framework which accounts for this characteristic. Hamalainen et al. (2010) review the case for mandatory SD policy for companies in the UK. Conversely to the findings of Caldwell et al. (2005), their findings suggest, that in the highly concentrated banking sector in the UK, only the six largest banks should be obliged to issue SD.

Chen and Hasan (2011) point out a side-concern which has not been discussed so far. They describe the possible scenario of investors 'colluding' with the bank (violating condition C3), in which the bank indirectly 'pays' investors in terms of higher coupons for not being monitored by them. This limitation to the efficiency of markets would require interest ceilings similar to those proposed by Calomiris (1999).

In a recent paper, Evanoff et al. (2011) discuss questions on the frequency and amount issued, again. They ask under which conditions SD spreads would be more inclined to be sensitive to bank risk and test their hypothesis that market depth is crucial here. Deeper markets, measured in terms of liquidity and size, could positively moderate the sensitivity since more information is provided around the issuance of debt instruments (condition C1). Risky banks could not avoid market exposure and market discipline anymore and low-risk banks could be incentivized to be transparent and benefit from lower funding costs (Evanoff et al., 2011).

Following the positive assessment of the effectiveness of SD in Evanoff et al. (2011), regulators should consider incorporating convertible features to SD into their reforms. Contingent convertible SD could increase the incentive for monitoring and lever risk sensitivity since investors would fear losses from conversion (Posner, 2010). These so-called CoCo

bonds received broad attention by researchers and regulators (Avdjiev et al., 2013) and have developed a relevant market since their set-up in 2009, when regulators forced banks to issue this specific type of SD. Basically, CoCo bonds are hybrid debt instruments which are converted into equity when a predefined trigger is breached. Their qualities as (subordinated) mezzanine debt make them first choice in overcoming classic issues of debt-equity agency problems. Further papers discuss the design of such contingent convertible requirements (Herring and Calomiris, 2011). In a more recent paper, Hilscher and Raviv (2014) present their research on how CoCo bonds influence risk-taking behavior and default probabilities of banks and find that optimally constructed CoCo bonds strongly impact these factors. They point out that contract terms (e.g. conversion ratio and thresholds) significantly matter here.

The discussed proposals could be adopted along current regulation on bail-in capital in Europe but require further research on their implementation. Future role of market discipline by SD will be determined by viable approaches and subject to regulators' decisions.

### 3. Empirical evidence on the monitoring phase

In the previous section 2, I laid out various conditions for verifying the existence and effectiveness of MD through SD and presented main proposals for implementation prevalent in the literature. The following part will revise empirical findings from prior studies and offer an own empirical approach to the assessment of risk sensitivity of SD. I will test for the monitoring phase of MD, more specifically for condition 3 in the framework in Figure 1 (vide supra). A positive assessment of the awareness of bank's risk by investors would partially prove MD.

#### 3.1. Prior empirical studies

First studies, published in the 1980s as a reaction to higher perceived risk in the US banking sector, show ambiguous results for the significance of efficient monitoring. For a sample period prior to 1991, Avery et al. (1988) do not find significant relationship between SD spreads and bank's risk-taking behavior. Later studies reviewing their findings explain the lack of explanatory power with implicit government guarantees in place (violating condition C2), possibly distorting market prices. Bail-outs of banks like Continental Illinois were then common practice by the Fed. Another study by Gorton and Santomero (1990) extends the previous model, adjusting it for leverage ratios which were not considered earlier. The rationale behind this adjustment comes from the insight that SD rather behaves like equity and higher risk could lead to higher returns. The findings of Avery et al. (1988) are confirmed, refusing the hypothesis that SD spreads and investors incorporate risk measures.

The second generation of studies, emerging in the 1990s, uses more recent data and comes to a positive assessment on the existence of SD sensitivity to risk measures. For this

period, there was a structural change in the banking landscape due to the Federal Deposit Insurance Corporation Improvement Act ("FDICIA") in 1991, when the US government ended their bail-out policy of banks considered TBTF. Flannery and Sorescu (1996) provide this opposing evidence for the 1983-1991 period, finding a significant correlation of risk measures (especially on credit ratings) for bank holding companies for later years, when banks were not considered TBTF anymore.

Most studies, until the turn of the millennium, focused on US markets and banks. Sironi (2003) presents a first study on European banks, methodologically similar to previous studies on US banks, with a sample period from 1991 to 2000, also confirming a sensitive relationship. His findings give evidence that (1) investors are sensitive to bank risk, using stand-alone credit ratings which exclude the impact of implicit guarantees, (2) sensitivity has increased with disappearing guarantees from governments and (3) public banks (especially referring to Germany) show no spread-relationship of risk measures since they enjoy explicit government guarantees. de Mendonça and Loures (2009) present a first study for an emerging country, Brazil, but only find weak evidence of MD. These studies aside, the number of recent studies using panel regressions, like I do, is limited and therefore updates are overdue.

There remain important caveats to cross-sectional studies investigating yield-risk relations (Bliss et al., 2001). Besides giving ambiguous evidence, most of the studies do not analyze the relevance of timeliness. Moreover, while most of the studies use accounting variables or credit ratings to explain different spread levels and movements, only few like Park and Peristiani (1998) make use of variables on market condition.

### 3.2. The market for subordinated debt in Europe

Using available data from Bloomberg data services, I will briefly describe the primary market for SD in Europe, based on own calculations. The total sample consists of 3712 bond issuances during the period from 1982 to 2017, whereas I will focus on the post-crisis period from 2007 to 2016 in the analysis.

Subordinated bonds are predominately denominated in EUR (65% of all issuances), followed by denominations in USD (17%), GBP (8%) and other currencies (10%). European banks have issued a total of EUR 326bn, USD 204bn and GBP 45bn over the analyzed period. The largest Euro issuers (share of all issuances: Italy 24%, Spain – 21%, the Netherlands – 9%, France – 8% and Great Britain – 7%) are accounting for EUR 224bn or 69% of the total amount issued in Euro.

Analyzing the amount issued over time, I observe an increasing amount of Euro issuances from EUR 23.8bn in 2000 peaking at EUR 52.5bn issued in 2007, but decreasing again to EUR 19.9bn in 2016. The same pattern is observed for USD issuances (USD 6.3bn in 2000, USD 24.1bn in 2009 and USD 21.8bn in 2016). Total market size of SD seems to be at moderate to low levels at this point of time.

With an average of 10 years to maturity, 42% of all issuances are maturing at 10 years from issuance and 20% are perpetual bonds. Half of all bonds are callable, containing a premium for the call option, while only 43% come with no further specification on maturity type. Average spread-to-benchmark of the debt notes at issuance in my sample is 336 BPS (with a standard error of 186 BPS) and coupons are oftentimes paid on a floating basis. Average credit ratings of these high-yield instruments are ranging from BBB (Standard&Poor's and Fitch) to Baa3 (Moody's).

Secondary markets for SD, however, are still considered relatively illiquid in Europe, as I infer from the fact that subordinated bond prices are only rarely quoted by brokers.

### 3.3. Hypotheses

I will narrow the focus of my empirical work to testing the necessary condition for MD, the monitoring phase, and want to assess the risk perception and sensitivity of SD investors. This is to determine which accounting variables are significantly contributing to the spread of SD bonds. Finding significant evidence for risk-indicating variables explaining spreads would substantiate market monitoring by investors as discussed in section 2.4.1. The view that market efficiency is per se given should not be considered as self-evident in the banking sector according to Flannery (2001), especially when governments make use of guarantees, as discussed in section 2. The hypotheses reflecting my approach are formulated in the following way:

Hypothesis 1 (H1): Spreads of subordinated bonds are significantly higher than spreads of senior bonds.

Hypothesis 2 (H2): External credit ratings have significant influence on SD spreads, investors perceive the risk of default assessed by external rating agencies.

Hypothesis 3 (H3): Key accounting ratios have significant influence on SD spreads. Investors perceive and price key risk derived from leverage, profitability, asset quality and liquidity.

Whereas H1 and H2 can be tested with decent effort, H3 requires more intense dedication to data collection and modelling and constitutes my main work. It should be noted that H1 is rather an auxiliary test, necessary because my further regressions only test subordinated bond spreads solely and not relative to senior debt. A comparison of both groups of debt would be feasible but harder to implement, given a lack of data and time constraints. If I accept H1, as I expect per definition, it can be inferred that subordinated bonds are high-yield bonds. This potentially makes them more sensitive to risk and SD is thus an adequate instrument for MD. H2 and H3 will then be testing spread sensitivity of subordinated bonds towards either ratings or accounting measures.

Furthermore, it is important to note that my analysis will not be able to render a differentiated answer to the question

how higher spreads influence a bank's reaction on asset composition and risk management decisions. To test for this type of market influence, one can inverse the spread relation and regress asset composition on spreads (Nguyen, 2013).

### 3.4. Methodology and data

Regarding H1, Welch's t-test is applied to test the null-hypothesis that the group of senior bonds and subordinated bonds have equal means. Welch's t-test is used instead of Student's t-test due to the different sample sizes across each rating group and unequal variances across samples. Filtered for available spread data, my analyzed sample includes 680 subordinated and senior bonds issued from 2007 to 2017. Data is retrieved from Thomson One/SDC Platinum by filtering the payment rank for subordinated bonds (excluding flags containing 'senior', such as 'subordinated senior bonds') and senior bonds (excluding flags containing 'subordinated', respectively). Cardinalized Moody's ratings (see Table X in the appendix) are normalized to a scale from 1 to 10 and thereby clustered in groups to facilitate interpretation. To be able to compare groups of homogeneous bonds, I test for the difference in each rating class. Only classes 1 to 6 are comparable here, lower rating classes are only populated by subordinated bonds or do not present sufficient observations.

To examine the influence of ratings (H2), I will run a simple OLS regression from equation (1), my dependent variables being Moody's and Standard & Poor's rating.

$$SPREAD_j = f(\text{S\&P Rankings}_j \vee \text{Moody's Ratings}_j) \quad (1)$$

A panel regression with ratings would require time-series data instead of static ratings only available. Furthermore, ratings are only adjusted annually (if ever) and changes seem to be too small to be relevant, which makes them a less suitable variable for a panel regression. Filtered for availability of spread data and of ratings, the sample analyzed includes 360 subordinated bonds issued from 2007 to 2017. Ratings, also retrieved from Thomson One/SDC Platinum and Bloomberg, are cardinalized (see Table X in the appendix) and normalized to a scale from 1 to 10 to facilitate comparison between rating agencies.

Regarding H3, testing the sensitivity of SD spreads towards risk factors, a wide variety of studies (see section 3.1) relies on cross-sectional time series models (panel regressions). Studies using this model are evaluating how risk measures of a bank, regarding the quality of the loan portfolio and other accounting data (e.g. capital ratios, leverage, profitability, etc.), can explain the risk premium to assess to which extent SD investors are risk-sensitive to available information. My methodology follows the well-accepted and widespread approach presented in Sironi (2003). I will give an updated review of his evidence from 2003 for European banks and use a wider set of variables to explain the spread.

The model is specified as follows with each observation  $i$  for a specific bank and year:

$$SPREAD_j = f(RISK_j) + \epsilon_j \quad (2)$$

RISK is a measure of (default) risk which should be perceived by investors depending on (book) accounting values.

$$RISK_j = f(\text{Leverage}_j, \text{Profitability}_j, \text{Asset Quality}_j, \text{Liquidity}_j) \quad (3)$$

For the dependent variable (SPREAD), three different specifications apply:

SPREAD (prim): These are primary market spreads (to benchmark), extracted as transaction information from all available deals from Bloomberg and Thomson One/SDC Platinum. Static spreads at issuance are aggregated for each year and bank to allow for a cross-sectional analysis. As actual transaction prices (instead of quotes by brokers on secondary markets), they provide a genuine view of the bank's cost and are better observable than secondary market spreads.

SPREAD (sec): I collected daily secondary market spreads of about 700 traded instruments (the other 1000 bonds not quoting prices) from Bloomberg and aggregated them on an annual basis for each bank.<sup>5</sup> Erroneous observations due to wrong quotation (e.g. extremely large spreads) and negative spreads are cleaned from the dataset.

$\Delta$ SPREAD (sec): Calculated from SPREAD (sec) as the approximated relative annual change for year  $t$ , which should provide results independent of the spread level.

$$\Delta SPREAD(sec) = \ln\left(\frac{SPREAD(t)}{SPREAD(t-1)}\right) * 100 \quad (4)$$

For the independent variables, serving as risk measures (RISK), I started testing the specification from Sironi (2003) but added own variables to test a wider variety of specifications. The presented specifications on the choice of independent variables result from iterative approaches maximizing the significance level for each variable. The bank-specific accounting data stem from Thomson Reuters DataStream. Independent variables are standardized to their standard error. Being aware of the conditions and caveats of such standardization, I will apply it for better interpretation of the magnitude.<sup>6</sup> The sample is sufficiently large and my explanatory variables are approximately normally distributed. A description of all variables is presented in Table IV in the appendix, together with Table V explaining the expected sign of the explanatory variables chosen. For a sample summary and correlation matrix, see Table VI and VII in the appendix. From the latter, I conclude that correlation within the sample is reasonable and does not distort my regression to great extent.

<sup>5</sup>The first attempt to gather traded spreads focused on meticulous own work compiling a list of representative and comparable individual subordinated bonds in terms of amount issued, currency and maturity date. However, I found that this sample is not representative and only found weak significance in later regressions.

<sup>6</sup>Standardization generally requires sufficient sample size, normal distribution of variables and ignores different standard errors between clusters (here: banks).

The Hausman test suggests fixed effects as the most efficient model for my panel regression. A test indicating heteroscedasticity for my data leads me to use robust standard errors in the model. Furthermore, after running an adequate statistical test, time-fixed effects are implemented as an alternative specification, which render more significant results and notably higher explanatory power. F-tests are used to determine if the coefficients are jointly different from zero.

My panel covers the period from 2007 to 2016 for a sample of 30 large EU banks and 2 large Swiss banks. Size is being measured in terms of total assets; the sample further includes the largest issuers of SD not present in the list of largest banks. See Table IX in the appendix for the list of sample banks.

The impact of sample selection bias for the sample period and entities should be limited due to various reasons: First, a surge in issuances already began around the years around 2005/2006 due to new Basel II regulation emphasizing MD. Second, big mergers across my sample banks were not frequent. Third, the presence of survivorship bias should be limited. The constituent's list of the largest European banks did not vary significantly, except for German public banks (various 'Landesbanken' and DZ Bank), vanishing from the list. Since those banks enjoy explicit guarantees, their exclusion does not harm my results for the sample of private banks.

### 3.5. Results and implications

Regarding H1, I find that subordinated debt (in contrast to senior debt) exhibits a significantly higher spread level, ranging from 50 to 120 basis points ("BPS") and a higher variation of spread level by looking at the standard errors. See Table 1 on the next page for the overall results of Welch's t-test. SD can thus be an adequate measure for exercising MD since investors are exposed to generally higher risk (measured by spreads).

From Table 1, one might observe that spread levels are generally rising along higher rating classes, however, I did not test for significance of this relationship so far. The test for H2 in Table 2 on the next page delivers results to this question. I can accept the hypothesis that credit ratings have significant impact on primary market spreads. Spreads for bonds with a lower/worse credit rating class increase significantly for each rating class.<sup>7</sup> For S&P ratings, starting from the non-significant intercept close to 0 BPS, spreads significantly increase by approximately 80 BPS per rating class. For Moody's ratings, starting from the significant intercept at 46 BPS, spreads significantly increase by approximately 60 BPS per rating class. I conclude that rating agencies play a vital role in providing information to SD holders in primary markets. There exists a clear and sensitive relationship between rating classes and spread level.

Interpreting the panel regression linked to H3, I find that market participants are well-aware of different risk profiles

across banks over time and price this observed risk on primary and secondary markets for SD bonds. Results presented in Table 3 (and Table VIII in the appendix for a model without time-fixed effects) can be read in the following way: A change of the explanatory variable by one standard deviation (due to standardization) leads to an increase or decrease in spread levels measured in BPS (SPREAD) or to an increase or decrease in the relative change of spread measured in percentage points ( $\Delta$ SPREAD).

Summarizing the most relevant results of the time-fixed effects regression in Table 3, based on significance levels of 1% or 5% and on high magnitude, I find the following: Specification (1) in Table 3 reports the spread level over time, peaking in 2012. For specifications (2) to (4), with a change of standard error by one unit, return on assets decreases spreads by 120 BPS. Equity to capital increases spreads by 225 BPS, loan loss reserves decrease spread by 360 to 620 BPS and provision for loan losses increases spreads by 130 to 225 BPS. Loan losses to loan loss reserves decrease spreads by 110 to 200 BPS. Non-performing loans to equity increase spreads by 400 to 715 BPS. Total loans to total capital unexpectedly reduce spreads by 160 to 175 BPS. Interest coverage ratio increases spreads by 60 BPS. The change in spreads ( $\Delta$ SPREAD) from specification (5) decreases unexpectedly with non-performing loans to total loans by 15 percentage points ("PPS").

Without time-fixed effects, I observe loss in explanatory power, reflected in lower levels of the coefficient of determination ( $R^2$ ). These results are therefore only reported in the appendix in Table VIII. I then find that (with a change in standard error by one unit) return on assets decreases spreads by 90 to 200 BPS, leverage decreases spreads by 225 to 290 BPS, equity to capital decreases spreads by 160 BPS and loan loss coverage increases spreads by 40 to 45 BPS, just as provision for loan losses increases spreads by 155 BPS. Non-performing loans to loan loss reserves increase spread by 150 to 170 BPS. Cash and securities to current debt portion decrease spread by 90 to 170 BPS.  $\Delta$ SPREAD rises for a change in standard error by one unit with return on assets by 23 PPS, with equity to capital by 14 PPS, with provision for loan losses by 28 PPS, with non-performing assets to total assets by 150 PPS. It falls with loan loss reserves by 85 PPS, with non-performing loans to total loans by 36 PPS (unexpectedly) and with total loans to total assets by 105 PPS (also unexpectedly).  $\Delta$ SPREAD also rises with the ratio of cash and securities to total deposits by 65 PPS. The inverse and overly high magnitude of non-performing loans to total loans and to equity should be omitted due to high correlation of 0.85. One should notice, however, that the explanatory power measured by  $R^2$  is very low for the specification of the model with  $\Delta$ SPREAD.

Based on this regression, I can say that many accounting variables are found to be with significant influence and magnitude to the spread level or the relative change in spread level. Some interaction variables from previous studies (Sironi, 2003) linked to leverage, ROALEV and LLRLEV, do not matter as risk variables to spreads in my regressions.

<sup>7</sup>Cardinalized and normalized on a scale from 1 to 10 with 10 being the worst rating class.



**Table 1:** Welch's t-test for difference between subordinated and senior bond spreads

Table 1 reports the results of Welch's t-test, testing for a significant difference in means of spread levels between subordinated and senior bonds. Dependent variable SPREAD (prim) are spreads to benchmark at issuance. The bonds are clustered in six rating classes obtained from the cardinalized Moody's ratings of the issued instrument. Moody's ratings are chosen because the border to non-investment grade lies just between rating class 5 and 6 which does not distort the means; Standard errors in parentheses; Significance level for difference: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

| Rating Class | SPREAD (prim)       |                     |                     |                   |                     |                  |
|--------------|---------------------|---------------------|---------------------|-------------------|---------------------|------------------|
|              | 1                   | 2                   | 3                   | 4                 | 5                   | 6                |
| Bond Group   |                     |                     |                     |                   |                     |                  |
| Senior       | 41.46<br>-7.578     | 130.5<br>-8.437     | 95.93<br>-4.409     | 193.1<br>(25.75)  | 249.4<br>(18.78)    | 417.0<br>(37.61) |
| Subordinated | 162.4<br>(22.50)    | 180.6<br>(12.94)    | 211.5<br>(16.57)    | 239.0<br>(17.12)  | 311.1<br>(16.01)    | 407.6<br>(24.02) |
| Combined     | 50.47<br>-7.901     | 155.0<br>-7.850     | 142.6<br>-8.446     | 225.7<br>(14.36)  | 291.3<br>(12.71)    | 409.4<br>(20.60) |
| Difference   | 120.9***<br>(23.75) | 50.13***<br>(15.45) | 115.6***<br>(17.15) | 45.92*<br>(30.92) | 61.76***<br>(24.68) | 9.355<br>(44.63) |
| N (Senior)   | 87                  | 101                 | 96                  | 25                | 33                  | 7                |
| N (Subord.)  | 7                   | 97                  | 65                  | 61                | 70                  | 31               |
| N (total)    | 94                  | 198                 | 161                 | 86                | 103                 | 38               |

**Table 2:** Standard OLS regression of spread on credit ratings

In Table 2, I test for the relationship between credit ratings and the spread to benchmark at issuance for subordinated bonds, SPREAD (prim), in two different OLS regressions for each rating agency. Ratings are cardinalized (see Table X in the appendix) and normalized on a scale from 1 to 10 to allow comparison between the two different rating agencies; Standard errors in parentheses; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ ;

| Variables      | SPREAD (prim)          |                     |
|----------------|------------------------|---------------------|
|                | Standard&Poor's<br>(1) | Moody's<br>(2)      |
| Rating         | 83.93***<br>(6.824)    | 58.78***<br>(4.271) |
| Constant       | 0.309<br>(22.26)       | 45.91**<br>(17.74)  |
| N              | 360                    | 355                 |
| R <sup>2</sup> | 0.297                  | 0.349               |

This insignificant relationship might be explained by the fact that I use book values instead of market values for equity and debt. Some results are unexpected, conflicting with the expectations from Table V in the appendix and should receive more attention in future research.

Implications from my results are that investors can discriminate between different risk levels in banks and are especially sensitive to return on assets, leverage, asset quality (non-performing loans to equity), provision for loan losses and loan loss reserves. Despite high significance of my findings, the standard errors are relevant in size, meaning that the effects can differ between banks. Furthermore, primary market spreads seem to be with higher significance compared to secondary market spreads, other specifications of SPREAD (sec) are not reported due to low significance.

Beyond the scope and beyond natural constraints my thesis is subject to, it would be feasible and interesting to analyze some aspects in more detail. My methodology could be extended to cover a larger sample and a longer time hori-

zon. Furthermore, one could use market values for some variables such as leverage, adjust time intervals (quarterly data, if available) and lag these explanatory variables to investigate timeliness of the relation. Considering the specification of my model, further extensions could include aggregated ratings as a time-series and any other variables that could serve as risk indicators. To attain an optimal model, one should systematically choose variables, for example by applying Davidson-Mackinnon P-test. Furthermore, I would advise testing the difference between small and big banks (based on total assets) to uncover implicit guarantees.

#### 4. Conclusion

In this thesis, I reviewed the concept of market discipline exemplified by subordinated debt in the context of European regulation after the financial crisis.

By analyzing the channels of market discipline, I highlighted the relevant conditions for effective market monitor-



**Table 3:** Linear, time-fixed effects panel regressions of spreads on accounting variables

Table 3 reports the OLS regression results of my panel over the sample period from 2007 to 2016. Dependent variable is the mean of primary and secondary market SPREAD (measured in BPS) and change in SPREAD ( $\Delta$ SPREAD, measured in percentage points) for each SD-issuing bank for each calendar year. Explanatory variables are reported in Table IV and summarized statistically in Table VI. Independent variables are standardized to their standard error. Estimation method is “fixed effects” augmented with time-fixed effects. Standard errors are set to be robust. Observations with high magnitude and significance at 5% level are highlighted in bold. F-test statistic for common significance of coefficients is not reported, however, the test is significant for every specification at the 1% level.; Standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1;

| Variables      | SPREAD (sec)        | SPREAD (prim)        |                      | $\Delta$ SPREAD (sec) |                      |
|----------------|---------------------|----------------------|----------------------|-----------------------|----------------------|
|                | (1)                 | (2)                  | (3)                  | (4)                   | (5)                  |
| ROA            |                     | -116.4**<br>(41.29)  | -124.5**<br>(47.77)  | -118.6**<br>(47.00)   | 9.640<br>(7.058)     |
| LEV            |                     | -84.60*<br>(41.16)   |                      |                       |                      |
| EQ             |                     |                      | 231.9***<br>(39.58)  | 223.8***<br>(48.78)   | 9.669*<br>(4.901)    |
| LLR            |                     | -365.9***<br>(124.5) | -358.3***<br>(105.3) | -617.3***<br>(184.7)  |                      |
| PROVLL         |                     | 132.6**<br>(58.41)   | 201.2***<br>(64.32)  | 225.3***<br>(57.27)   | 18.64*<br>(9.620)    |
| LLtoLLR        |                     | -118.9***<br>(25.67) | -112.2***<br>(19.93) | -203.4***<br>(63.39)  | -7.671*<br>(3.986)   |
| NETLL          |                     |                      |                      | 89.57<br>(62.27)      |                      |
| NPLtoE         |                     | 443.4***<br>(134.7)  | 403.4***<br>(117.5)  | 715.5***<br>(169.7)   |                      |
| NPLtoTLO       |                     |                      |                      |                       | -15.81***<br>(5.182) |
| TLOtoTCAP      |                     | -40.67<br>(23.60)    | -159.9***<br>(26.62) | -175.5***<br>(34.84)  |                      |
| INTCOV         |                     |                      |                      | 58.73***<br>(15.43)   |                      |
| 2008           | 146.0***<br>(14.47) | 5.660<br>(77.95)     | 21.09<br>(77.89)     | -17.50<br>(77.95)     |                      |
| 2009           | 363.9***<br>(35.24) | 207.3***<br>(66.52)  | 95.54<br>(109.0)     | 82.19<br>(108.4)      | -49.76***<br>(9.356) |
| 2010           | 181.0***<br>(29.69) | -40.39<br>(109.4)    | -150.2<br>(101.1)    | -141.8<br>(116.4)     | -150.7***<br>(11.50) |
| 2011           | 519.5***<br>(88.54) | 223.4<br>(166.7)     | 121.3<br>(171.6)     | 173.6<br>(142.3)      | -24.38**<br>(9.383)  |
| 2012           | 736.9***<br>(107.9) | 287.0**<br>(114.9)   | 226.9**<br>(100.9)   | 277.7**<br>(112.6)    | -68.86***<br>(11.35) |
| 2013           | 358.2***<br>(44.28) | 126.6<br>(100.6)     | 58.42<br>(82.63)     | 49.67<br>(90.65)      | -125.5***<br>(10.79) |
| 2014           | 176.6***<br>(22.19) | -109.9<br>(87.27)    | -176.1**<br>(67.10)  | -152.1**<br>(64.92)   | -122.7***<br>(14.55) |
| 2015           | 206.5***<br>(32.14) | -206.5*<br>(110.0)   | -310.7***<br>(89.51) | -317.8***<br>(81.71)  | -70.73***<br>(15.15) |
| 2016           | 260.8***<br>(31.50) | -192.1*<br>(102.8)   | -242.7**<br>(87.37)  | -293.8***<br>(79.46)  | -77.80***<br>(17.09) |
| Constant       | 105.3***<br>(28.16) | 448.8***<br>(72.28)  | 491.1***<br>(59.31)  | 517.8***<br>(65.12)   | 89.32***<br>(6.344)  |
| N              | 265                 | 64                   | 64                   | 63                    | 144                  |
| R <sup>2</sup> | 0.366               | 0.340                | 0.201                | 0.165                 | 0.665                |

ing and market influence, in brief: sufficient transparency and disclosure, no guarantees for bail-out and efficient capital markets. I gave an overview over the conditions for subsequent market influence, namely credible action (indirect influence), sizeable cost effects (direct influence) and adequate incentive structures. My literature review points out that spreads of subordinated bonds can possibly reflect the risk profile of banks and be implemented on a mandatory level along current regulation.

Complementing the current literature on the monitoring phase of market discipline, my empirical results contribute to the discussion about the sensitivity of subordinated debt instruments to bank risk. The main conclusions emerging from my own empirical work are as follows: Firstly, subordinated debt exhibits a significantly higher spread level of 50 to 120 BPS in contrast to senior debt and is potentially an adequate instrument for MD. Secondly, rating agencies play a vital role in providing information to subordinated debt investors in primary markets and there exists a highly sensitive relationship between an instrument's rating class and its spread level, ranging from 60 to 80 BPS per rating class (for ratings up to class 6 on a scale from 1 to 10). Thirdly, market participants are well-aware of different risk profiles across banks and price this difference on primary and secondary debt markets for subordinated bonds. Many accounting variables are found to be with significant influence and high magnitude to the spread level or the change in spread level. Spreads are significantly sensitive to ROA, equity to capital, loan loss reserves, provision for loan losses and non-performing loans; each to notably more than 100 BPS per standard error of change in the explanatory variable.

These insights have important implications regarding the future use of subordinated debt as a regulatory measure in Europe. Yield spreads might serve well as a market based risk indicator, supplementing book ratios and ratings, and could influence banks' risk decisions directly and indirectly. Given the positive findings on the monitoring phase of subordinated debt, most of the recent literature points towards the common sense that regulators should perceive the potential of subordinated debt and actively incorporate it into their framework.

The European market for subordinated debt is already sizeable with total issuances of EUR 326bn, USD 204bn and GBP 45bn in the years from 2007 to 2016. Although declining lately, it could be growing substantially if mandatory proposals are implemented by regulators. Such mandatory subordinated debt proposals are widely discussed in research, without coming to implementation in national or supranational frameworks yet. Future role of market discipline by subordinated debt will be determined by viable approaches and subject to regulators' decisions. From a European, but also from a global perspective, a common framework seems unavoidable and common standards indicating the payment rank would be desired to raise transparency.

Open questions emerging from my thesis provide avenue for further research. It would be crucial to reassess whether there is an optimal (minimum) level for shares of subordi-

nated debt. Future research might then wish to explain how the feedback of spreads influences bank managers and risk-taking exactly in extended moral hazard models. We still do not fully understand how banks incorporate higher spreads on subordinated debt markets into their behavior and how TBTF guarantees interfere with a reaction. Therefore, the role of implicit TBTF guarantees (or 'no bail-out' clauses vice versa) should be further discussed in their effect on subordinated debt spreads. This becomes especially crucial in the European Banking Union when member countries are deviating from commitments.

Concluding, I can say that the question motivating this thesis was convincingly answered with sufficient evidence for market discipline regarding the monitoring phase. Future regulation on bail-in capital in Europe should seek to adopt subordinated debt into their framework.

## References

- Admati, A. R. and Hellwig, M. F. Does debt discipline bankers? an academic myth about bank indebtedness, 1-37. Technical report, 2013.
- Avdjiev, S., Kartasheva, A., and Bogdanova, B. Cocos: a primer1. *BIS Quarterly Review*, pages 43–56, 2013.
- Avery, R. B., Belton, T. M., and Goldberg, M. A. Market discipline in regulating bank risk: New evidence from the capital markets. *Journal of Money, Credit and Banking*, 20(4):597–610, 1988.
- Barth, J. R., Caprio, G., and Levine, R. Bank regulation and supervision: what works best? *Journal of Financial Intermediation*, 13(2):205–248, 2004.
- Barth, J. R. and Wihlborg, C. Too big to fail and too big to save: dilemmas for banking reform. *National Institute Economic Review*, 235(1):R27–R39, 2016.
- Bernanke, B. S. Financial regulation and the invisible hand speech. Technical report, New York City Law School, 2007.
- Birchler, U. W. and Facchinetti, M. Self-destroying prophecies? the endogeneity pitfall in using market signals as triggers for prompt corrective action. *BIS Research Task Force Paper*, March, 2007.
- Bliss, R. R. and Flannery, M. J. Market discipline in the governance of us bank holding companies: Monitoring vs. influencing. *Review of Finance*, 6(3):361–396, 2002.
- Bliss, R. R. et al. Market discipline and subordinated debt: A review of some salient issues. *Economic Perspectives-Federal Reserve Bank of Chicago*, 25(1):24–45, 2001.
- Buck, T. Santander takes over 'failing' rival Banco Popular after EU steps in. *Financial Times*, 2017. URL <https://www.ft.com/>. June 2017.
- Caldwell, G. et al. *Subordinated debt and market discipline in Canada*. Bank of Canada, 2005.
- Calomiris, C. W. Building an incentive-compatible safety net. *Journal of Banking & Finance*, 23(10):1499–1519, 1999.
- Chen, Y. and Hasan, I. Subordinated debt, market discipline, and bank risk. *Journal of Money, Credit and Banking*, 43(6):1043–1072, 2011.
- de Mendonça, H. F. and Loures, R. F. V. Market discipline in the brazilian banking industry: an analysis for the subordinated debt holders. *Journal of Regulatory Economics*, 36(3):286–307, 2009.
- Demirgüç-Kunt, A. and Huizinga, H. Market discipline and deposit insurance. *Journal of Monetary Economics*, 51(2):375–399, 2004.
- Dewatripont, M. European banking: Bailout, bail-in and state aid control. *International Journal of Industrial Organization*, 34:37–43, 2014.
- Evanoff, D. D. and Wall, L. D. Subordinated debt and bank capital reform. *Working Paper*, Federal Reserve Bank of Atlanta, 2000.
- Evanoff, D. D., Jagtiani, J. A., and Nakata, T. Enhancing market discipline in banking: The role of subordinated debt in financial regulatory reform. *Journal of Economics and Business*, 63(1):1–22, 2011.
- Flannery, M. J. The faces of "market discipline". *Journal of Financial Services Research*, 20(2-3):107–119, 2001.
- Flannery, M. J. and Sorescu, S. M. Evidence of bank market discipline in subordinated debenture yields: 1983–1991. *The Journal of Finance*, 51(4):1347–1377, 1996.
- French, K., Baily, M., Campbell, J., Cochrane, J., Diamond, D., Duffie, D., Kashyap, A., Mishkin, F., Rajan, R., Scharfstein, D., et al. The squam lake report: fixing the financial system. *Journal of Applied Corporate Finance*, 22(3):8–21, 2010.
- Gorton, G. and Santomero, A. M. Market discipline and bank subordinated debt: Note. *Journal of money, credit and Banking*, 22(1):119–128, 1990.
- Goyal, V. K. Market discipline of bank risk: Evidence from subordinated debt contracts. *Journal of Financial Intermediation*, 14(3):318–350, 2005.
- Grossman, E. and Woll, C. Saving the banks: The political economy of bailouts. *Comparative Political Studies*, 47(4):574–600, 2014.
- Hakenes, H. and Schnabel, I. Banks without parachutes: Competitive effects of government bail-out policies. *Journal of Financial Stability*, 6(3):156–168, 2010.
- Hamalainen, P. Mandatory subordinated debt and the corporate governance of banks. *Corporate Governance: An International Review*, 12(1):93–106, 2004.
- Hamalainen, P., Hall, M., and Howcroft, B. A framework for market discipline in bank regulatory design (retraction of vol 32, pg 183, 2005). *Journal of Business Finance & Accounting*, 36(7-8):1039–1039, 2009.
- Hamalainen, P., Howcroft, B., and Hall, M. Should a mandatory subordinated debt policy be introduced in the united kingdom? evidence from the issuance activity of banks and building societies. *Contemporary Economic Policy*, 28(2):240–263, 2010.
- Hancock, D. and Birchler, U. What does the yield on subordinated bank debt measure? *SSRN eLibrary*, 2004.
- Herring, R. J. and Calomiris, C. W. Why and how to design a contingent convertible debt requirement. *Available at SSRN 1815406*, pages 1–58, 2011.
- Hilscher, J. and Raviv, A. Bank stability and market discipline: The effect of contingent capital on risk taking and default probability. *Journal of Corporate Finance*, 29:542–560, 2014.
- Hoang, K. T., Faff, R., and Haq, M. Market discipline and bank risk taking. *Australian Journal of Management*, 39(3):327–350, 2014.
- Hwang, Y.-S. and Min, H.-G. Market discipline of subordinated debt: Empirical evidence from japanese commercial banks. *Journal of Financial Risk Management*, 2(2):38–42, 2013.
- Jagtiani, J. and Lemieux, C. Market discipline prior to bank failure. *Journal of Economics and Business*, 53(2-3):313–324, 2001.
- Lall, R. From failure to failure: The politics of international banking regulation. *Review of International Political Economy*, 19(4):609–638, 2012.
- Lintner, P. Understanding Bank Recovery and Resolution in the EU: A Guidebook to the BRRD, 2017.
- Martino, E. Subordinated debt under bail-in threat. what is going to change in the debt market for bank industry? *University of Bologna Law Review*, pages 1–41, 2017.
- Mishkin, F. S., Stern, G., and Feldman, R. How big a problem is too big to fail? a review of gary stern and ron feldman's too big to fail: the hazards of bank bailouts. *Journal of Economic Literature*, 44(4):988–1004, 2006.
- Nguyen, T. The disciplinary effect of subordinated debt on bank risk taking. *Journal of empirical finance*, 23:117–141, 2013.
- Park, S. and Peristiani, S. Market discipline by thrift depositors. *Journal of Money, Credit and Banking*, pages 347–364, 1998.
- Posner, K. A. Thoughts on the squam lake report: Reengineering the financial system to better withstand extreme volatility. *Journal of Applied Corporate Finance*, 22(3):34–39, 2010.
- Sironi, A. Testing for market discipline in the european banking industry: evidence from subordinated debt issues. *Journal of Money, Credit and Banking*, pages 443–472, 2003.
- Van den Heuvel, S. J. The welfare cost of bank capital requirements. *Journal of Monetary Economics*, 55(2):298–320, 2008.



# The Impact of the Organic Label Halo Effect on Consumers' Quality Perceptions, Value-in-Use and Well-Being

Philipp Küst

*Freie Universität Berlin*

## Abstract

Food well-being is an innovative field of research analysing the complex consequences of food intake on body and mind. In face of mounting civilization diseases and environmental challenges promoting healthy and sustainable diets is crucial. For consumers it is difficult however, to assess the healthiness and environmental friendliness of a product. Food labels, like the organic one, are therefore used as extrinsic cues to help customers distinguish between alternatives.

This paper analyses how the organic label biases consumers' quality perception, a phenomenon that has been referred to as the organic label halo effect. It further intends to uncover the links between several quality dimensions and their consequences as reflected in value-in-use. Finally it aims to detect if those consequences ultimately lead to enhanced post-prandial well-being.

A quantitative study in an experimental canteen setting was conducted to answers those questions. Structural equation modelling (SEM) was applied to test the hypotheses. The results indicate a positive impact of the organic label halo effect on consumers' quality perceptions in terms of health & safety, environmental friendliness & animal welfare as well as prestige. Those in turn were shown to positively influence on several value-in-use dimensions, including social, altruistic, functional and hedonic value. Finally, the latter two were significantly related to well-being.

Hence, this research shows that providing organic food in a canteen ultimately enhances consumer well-being through inferential beliefs on quality and value evoked by the label. The findings help to better understand the links between food consumption and subjective well-being and are therefore of interest for policy makers and researchers around the world.

**Keywords:** food well-being; organic label halo effect; value-in-use; perceived quality; extrinsic cues.

## 1. Introduction

Prevailing global challenges like climate change or the proliferation of civilization diseases cause policy makers around the world to promote sustainable practices that contribute towards the collective well-being of societies (Larceneux et al., 2012). The conventional, industrialised agriculture and food production systems are often criticised for being co-responsible for many of these problems, for instance through their extensive use of pesticides and fertilizers (Lockie et al., 2004). Nowadays, a new generation of shoppers has emerged, who refuses the intensive agricultural practices and shows concern for animal welfare and the environmental impact of farming (Janet Eastwood, 1995). These customers desire more information on production related aspects so they can choose for themselves which product they want to purchase (Zanoli and Naspetti, 2002). This has been referred to as sustainable consumption behaviour.

The concept has been recognized by the United Nations as an important tool for environmental protection and social change and was included as one of the 17 UN Sustainable Development Goals that aim to transform the world until 2030 (Nations, 2015).

Food-labels are important tools to enable sustainable consumption behaviour as a means of information and transparency (Caswell and Mojduszka, 1996). They allow consumers to distinguish between products for example by signalling environmentally benign outcomes and therefore act as information tools (Grankvist et al., 2004). On the market, many different eco-labels exist, however, most of them are rather unknown to customers and hence also not trusted (Drexler and Fiala, 2018). The most well-known eco-label is the organic one. It dates from the 1970s in the US and has consistently been shown to be positively perceived by consumers (Howard and Allen, 2006).

Organic production started as an alternative movement in the 1920s in Europe, driven by the growing industrialization and agglomeration in the agricultural sector. Since then it has grown tremendously and is now regulated at national and supranational levels (Torjusen et al., 2001). As organic products are characterized by many so-called credence attributes that consumers cannot evaluate easily, like production methods, animal welfare or the use of pesticides and fertilizers, labels are used as extrinsic cues to indicate their hidden characteristics. However, since credence food labels are a relatively new phenomenon, the associated body of research is still rather limited (Fernqvist and Ekelund, 2014).

Existing research has focused mainly on identifying demographic characteristics of organic food consumers (e.g. Hughner et al., 2007; Lockie et al., 2004), their purchasing motivations (e.g. Magnusson et al., 2003; Zanolli and Naspetti, 2002) and relevant product attributes of organic produce (e.g. Harper and Makatouni, 2002; Wier et al., 2008). Some have also looked at the purchasing barriers including price, mistrust or convenience (e.g. Magnusson et al., 2001; Padel and Foster, 2005). A few authors have used means-end chain theory to identify the driving forces behind organic food consumption (e.g. Makatouni, 2002; Padel and Foster, 2005; Zanolli and Naspetti, 2002). This research follows the latter authors but aims to examine all constructs in the theory including attributes, consequences and end-goals and specifically the relationships between those concepts.

Attributes are the basis for quality evaluations but since they are not always objective, it is necessary to measure how customers subjectively perceive them (Lee and Yun, 2015). It has been postulated that consumers are often unaware of what the organic label actually certifies so they build their own associations and beliefs about its meaning (Vega-Zamora et al., 2014). These inferential beliefs are likely to positively influence their quality evaluations, a phenomenon that has been referred to as the organic label halo effect (e.g. Apaolaza et al., 2017; Lee et al., 2013). However, more research is needed in order to better understand the mechanisms of the effect in different settings (Apaolaza et al., 2018).

So far, only a few studies have linked organic product attributes to perceived value. Value-in-use is believed to stem from the consequences consumers associate with the use of a product (Woodruff, 1997). One study by Lee and Hwang (2016) found that eco-friendliness and food safety significantly impacted on perceived value. Furthermore, Vega-Zamora et al. (2014) identified that consumers experience hedonic pleasure when eating organic products. Research on those effects is however really scarce (Apaolaza et al., 2018). In fact, it is striking that many studies focusing on purchasing motivations for organic food completely miss out on the differentiation between purchaser and consumer and are therefore likely to be at best incomplete (Hughner et al., 2007).

Finally, there is only limited research on the underlying values of organic food purchasing (Padel and Foster, 2005). One end-goal that is commonly identified in those

studies is well-being (e.g. Padel and Foster, 2005; Zanolli and Naspetti, 2002). Indeed, in the last years, food well-being has emerged as an innovative research stream analysing the complex consequences of food consumption for the body and mind (Bublitz et al., 2013). In a recent contribution, Apaolaza et al. (2018) found a positive impact of the organic label halo effect on subjective well-being and asked for more research into that area. Moreover, in the latest *Ökobarometer* (2017), a representative study on organic food consumption in Germany, the participants referred to individual well-being as one of the top 5 reasons for buying organic products. The study further revealed favourable attitudes and a high demand for organic meals in canteens, where so far the supply is rather limited (*Ökobarometer*, 2017). Even though the well-being of the work-force is fostered by many companies (Diener, 2006), it is striking that so far, to the author's best knowledge, no research has yet analysed how organic food provision in canteens impacts on employee well-being.

This research intends to fill the identified literature gaps by evaluating the mechanisms of the organic label halo effect in a canteen setting. In accordance with means-end chain theory, it analyses how the quality attributes of food products relate to specific value-in-use dimensions in a consumption situation. Finally, this research contributes to the growing research stream of food well-being by assessing consumers' post-prandial well-being. Taken together, the following research questions are addressed:

1. How does the organic label halo effect impact on consumers' quality perceptions of food in a canteen setting?
2. What are the links between specific quality dimensions of organic food and perceived consequences as reflected in value-in-use?
3. Does providing organic food in canteens ultimately lead to enhanced post-prandial well-being through the benefits associated with its consumption?

To answer these questions, a quantitative study in a canteen setting was conducted. Participants were randomly assigned to two groups, one being treated with an organic food offer in a canteen setting. A t-test was used to prove the halo effect by analysing group differences in terms of respondents' quality evaluations. Subsequently, structural equation modelling (SEM) was applied to test for the hypothesized relationships between the constructs in the model.

The findings show a clear indication of a positive organic label halo effect on quality evaluations in a canteen. Moreover, the quality beliefs are significantly related to several value-in-use dimensions and ultimately well-being. Hence, this research argues that providing organic food in canteens positively biases consumers' quality perceptions leading to enhanced employee well-being through the benefits associated with its consumption. The findings of this research help to better understand how food influences consumer well-being and may therefore be useful for researchers as well as practitioners.



The remainder of the paper is structured as follows: In the literature review, the theoretical foundations of this research will be explained. Based on this analysis, hypotheses and a conceptual model will be derived. Thereafter, in the methodology chapter, the data collection and analysis methods will be justified. Having discussed the research method, the findings will be presented and subsequently discussed with reference to existing literature. The paper will end with a short summary, practical implications and limitations as well as recommendations for future research.

## 2. Literature Review

In the following chapters, a conceptual framework will be derived based on existing literature on organic food, value-in-use and well-being. First, means-end chains as the underlying theory of this research will be explained. Thereafter, the distinct constructs of the theory i.e. product attributes, consequences and ends will be discussed in detail.

### 2.1. Theoretical Foundation: Means-End Chains

Means-end chain theory (Gutman, 1982) relates products to personal values. It builds on previous work of researchers like Abbott (1955) who supposed that goods are not purchased and consumed for their own sake, but for what they mean and stand for and the experience they provide. In this theory, objects or products can be seen as means that help to achieve desired end-goals through the benefits of their consumption. In other words, the attributes of a good lead to specific consequences that contribute to higher-order ends (Grunert, 1995).

A great example of a means-end chain for a low fat product by Grunert (1995, p. 172) helps to differentiate between the constructs in the model and illustrates how they are related to one another: "[...] the concrete product characteristic low fat is linked to the abstract product characteristic fewer calories, linked to the consequences slimming (functional) and social acceptance (psychosocial), which leads to the values self-confidence (instrumental) and self-esteem (terminal)".

Similar chains are likely to occur in the case of organic food consumption. Zanolini and Naspetti (2002) for instance showed that the quality attributes of organic produce centring around ingredients, certified production methods and taste are related to perceived benefits when consuming and higher order ends like health and well-being. In fact, means-end chain models have frequently been used to identify the driving forces behind organic food consumption (e.g. Makatouni, 2002; Padel and Foster, 2005).

The following chapters are structured according to the components of means-end chains. First, it will be looked at how product attributes, including the organic label as an extrinsic cue, form the basis of consumers' quality perceptions. Thereafter, the discussion will turn towards the consequences of these attributes as reflected in value-in-use. Hypotheses will be formulated on how specific quality attributes

of organic food products influence on consumer value perceptions. Finally, the concept of well-being as an end-goal in organic food consumption will be presented.

### 2.2. Perceived Quality

This chapter takes a closer look at the concept of perceived quality. Based on cue theory, it will be argued that the label as an extrinsic attribute leads to the activation of descriptive and inferential beliefs creating a halo effect that biases consumers' quality evaluations. On the basis of existing literature, hypotheses will be derived on the effect of the label on specific quality attribute perceptions, including health & safety, environmental friendliness & animal welfare as well as prestige. To start with, the chapter will begin with a definition of quality and product attributes.

#### 2.2.1. Product Attributes as the Basis for Quality Perceptions

Product attributes form the basis of means-end chain models. Any product represents a bundle of attributes (Gutman, 1982). The term refers to the properties or characteristics of a product, i.e. what a good encompasses or is made of. Hence, product attributes simply constitute quality aspects of a good that influence consumers' quality evaluation (Grunert, 1995).

Perceived quality has been defined as "the customer's perception of the overall quality or superiority of a product or service with respect to its intended purpose, relative to alternatives" (Aaker, 1991, p. 85). According to Porter (1980), the superior quality of a product is the basis for consumer value and may lead to a sustainable competitive advantage. In the words of Golder et al. (2012, p. 1) "firms compete on quality, customers search for quality, and markets are transformed by quality".

Zeithaml (1988) ascertained the influence of product intrinsic and extrinsic attributes on quality perceptions. According to her, perceived quality is a global assessment of attributes on a higher level of abstraction that is similar to attitude, but differs from objective quality (Zeithaml, 1988). This idea is supported by the notion, that quality perceptions do not necessarily require experience but can just be formed based on specific attributes of a product (Rust et al., 1994).

In the literature, three different categories of product attributes are mentioned; search, experience and credence attributes (Darby and Karni, 1973; Nelson, 1970). Search attributes include characteristics that can be seen and assessed via the senses before purchase, like the colour or smell of a good. Experience attributes encompass things that can only be appraised during consumption like a food's taste, while credence attributes, like production methods, cannot be evaluated by customers even after consumption (Caswell and Mojduszka, 1996).

Many of the issues that are of increasing importance for consumers nowadays belong to the category of credence attributes for which they have to rely on information provided by others (Ophuis and Van Trijp, 1995). As customers cannot assess the safety or environmental friendliness of a product,

organic food is generally considered to be a so-called credence good (Caswell and Mojduszka, 1996). Hence, the only way to build quality evaluations and to distinguish organic from conventional products is to refer to a cue (Vega-Zamora et al., 2014).

### 2.2.2. Cue Theory

Cue Utilization Theory was mainly developed by Cox (1967) and Olson and Jacoby (1972). In order to choose between a given set of alternatives, customers try to overcome the existing information asymmetry and uncertainty by using quality indicators, so-called cues, that help them evaluate and choose between the offers (Cox, 1967). A cue is valued because consumers think it indicates a product's quality characteristics (Steenkamp, 1990). According to Cox (1967, p. 625), consumers "prefer cues highest in information value, that is, cues that best reduce the amount of their uncertainty." This is especially the case for goods with many credence attributes that cannot be determined by customers otherwise. It is important to mention that for Steenkamp (1990) quality cues classify as search attributes since they can easily be observed and analysed visually.

One differentiates between intrinsic and extrinsic product cues (Olson and Jacoby, 1972). The former ones include indicators such as shape or appearance that are related to the physical attributes and cannot be changed without altering the good itself. In contrast, extrinsic cues can be manipulated without changing the composition of the product for instance price, brand and information about manufacturing processes or animal welfare (Olson and Jacoby, 1972). Examples of extrinsic cues that help customers to assess the contents and production methods of food products are nutritional tables or eco-labels (Steenkamp, 1990).

### 2.2.3. Organic Label as a Cue

The organic label is a marketing tool (Drexler and Fiala, 2018). It helps customers to identify certain quality characteristics of goods that are otherwise invisible (Howard and Allen, 2006). In fact, the label might be the only reliable information consumers have on credence attributes such as production related issues. Hence, the label acts as a cue, transforming credence characteristics into search attributes (Caswell and Mojduszka, 1996).

In the European Union, common standards for organic produce have been defined that "aim at producing products of high quality; [...] that do not harm the environment, human health [...] or animal [...] welfare." (EC, 2007, p. L189/6). The regulation forbids genetically modified (GM) content and highly restricts the use of chemicals, fertilizers, pesticides and food additives. Furthermore, clear standards for animal rearing exist including housing conditions, breeding and feeding. The use of growth hormones for instance is strictly prohibited while access to open air areas is required by the regulation (EC, 2007).

Compliance with these standards is shown by using certified labels that are often directly printed on the packaging and prove a producer's attempts to ensure environmentally

sound production systems (Drexler and Fiala, 2018). The German national organic label ("Bio-Siegel") was introduced in 2001 for all products complying with the European regulations of organic farming. It may be used alongside the mandatory European logo that was introduced in 2010 to have a common label throughout the EU (BLE Bundesanstalt für Landwirtschaft und Ernährung, 2016).

Organic labels can be classified as positive labels, signifying better environmental performance in contrast to negative ones that warn about the harmful impacts of a product (Grankvist et al., 2004). For producers the certification is a way to distinct their offers in the market and achieve higher appreciation for their goods (Vega-Zamora et al., 2014). According to MINTEL (2000), the word organic has become a brand itself, thereby facilitating the charging of price premiums (Howard and Allen, 2006). In other words, the organic label acts as a quality enhancer and might lead to a competitive advantage for the supplier (Grankvist et al., 2004).

In consumer research, it is important to differentiate between the standards of labels as defined by regulators and the perceptions of customers who, according to Vega-Zamora et al. (2014), will have their own interpretation of the label. In fact, it has been postulated that even though most consumers have heard of the term organic and know some of its central features such as being chemical-free, they are often unaware of the actual standards of organic farming and production (e.g. Hill and Lynchehaun, 2002; Yiridoe et al., 2005). In other words, regardless of the official meaning of the label, customers will build their own associations and their subjective perception will impact on how they see the product (Vega-Zamora et al., 2014).

### 2.2.4. Organic Label Halo Effect

Based on spreading activation theory (Collins and Loftus, 1975), it can be reasoned that the meaning of terms such as natural or organic stems from an individual's inferences of these words (Larceneux et al., 2012). Consumers associate different aspects with organic labels, which can be classified as descriptive or inferential beliefs based on whether they result directly from the information provided by the label or arise indirectly through cognitive associations. Those inferential beliefs stem from inferences between a cue, specific product attributes and their perceived benefits and are especially prevalent in the case of credence characteristics, which cannot be easily validated by the customer (Steenkamp, 1990).

As argued before, quality perceptions are formed on the basis of specific product characteristics (Ophuis and Van Trijp, 1995). With regards to organics, the label acts as a heuristic cue indicating the perceived superiority of the good in comparison to others (Vega-Zamora et al., 2014). Such a perceptual bias, where one specific attribute, in this case the label, affects the evaluation of others and thus determines the overall assessment of an object, has been referred to as a so-called halo effect (Thorndike, 1920).

The existence of an organic label halo effect has been previously shown (e.g. Apaolaza et al., 2017; Lee et al., 2013).

In fact, organic food is generally believed to be of higher quality than conventional alternatives (e.g. Hill and Lynchehaun, 2002; Larceneux et al., 2012). In a study by Apaolaza et al. (2017) on organic wine, the label led to higher ratings in almost all of the sensory attributes measured like aroma, taste and healthiness extending towards a higher hedonic liking and purchase intention. Similar effects were found by Vega-Zamora et al. (2014). In their focus-group study, organic food was considered advantageous in most of the properties studied, including flavour and naturalness, while at the same time participants were mostly unaware of the differences between organic and conventional products. In their words, the latter ones are not bad, but the organic ones are just better (Vega-Zamora et al., 2014).

Hence, the label biases a product's evaluation in different quality dimensions (Apaolaza et al., 2017; Fernqvist and Ekelund, 2014). In the following, a closer look will be taken at three quality dimensions in particular: health & safety, environmental friendliness & animal welfare as well as prestige. The former two have been shown to be the most important quality characteristics of organic food (e.g. Magnusson et al., 2003; Makatouni, 2002; Wandel and Bugge, 1997). Prestige was included as an interesting dimension which has not yet received much of attention in the literature.

#### *Health & Safety*

Health related aspects can be considered main elements of perceived product quality (Vega-Zamora et al., 2014). They belong to the typical credence attributes of groceries (Ophuis and Van Trijp, 1995). A food of higher quality is commonly understood as being healthier and vice versa. Health & safety aspects of groceries centre around ingredients, nutrients, additives and chemical residues. The organic label thereby acts as a cue that leads to the activation of consumer beliefs about the healthiness of the products (Vega-Zamora et al., 2014). Following different authors (Lee and Hwang, 2016; Vega-Zamora et al., 2014), in this research, naturalness was included as a sub-category of health & safety. As consumers consider natural products to be unprocessed and contain no additives the link to health & safety aspects is rather strong.

In a computer-based experiment by Schuldt and Schwarz (2010), participants rated an organic cookie as having lower calories than the non-labelled version thereby proving the halo effect of the label on health related aspects. This has also been shown for other products such as yoghurt (Lee et al., 2013). Furthermore, some studies identified consumer beliefs of organic food being more nutritious (Hill and Lynchehaun, 2002; Hoefkens et al., 2009). Hoefkens et al. (2009) for example found that compared to conventional vegetables, organic produce is considered superior in terms of nutritional value. To date, no consensus exists among researchers whether organic food actually comprises more nutrients than conventional ones (e.g. Yiridoe et al., 2005). However, customers perceive them to have less calories and more nutrients and hence, as healthier than their conventional alternatives (Hoefkens et al., 2009; Magnusson et al., 2001).

Moreover, consumers believe organic food to be produced in a natural way, thereby containing less chemicals (Hoefkens et al., 2009; Padel and Foster, 2005). Organic products are seen to be free from added substances, artificial ingredients and genetically modified materials (Harper and Makatouni, 2002). They are regarded as authentic, real food that comes directly from the farm to the customer (Vega-Zamora et al., 2014). Consequently, organic food is also considered to be safer than conventional alternatives (e.g. Hoefkens et al., 2009; Wier et al., 2008). According to Lockie et al. (2002), food safety includes issues around chemical residues or GM-content. Lower residues are of particular importance for fruits and vegetables as they are often eaten with their skin (Padel and Foster, 2005). The cognitive link between organic and food safety is regularly reinforced by food scares and outbreaks of animal diseases with the most prominent example being the BSE crisis (Hughner et al., 2007). Accordingly, it is hypothesized that:

H 1.1: The organic label (no label) leads to higher (lower) quality evaluations in terms of health & safety.

#### *Environmental Friendliness & Animal Welfare*

The environmental quality of food is gaining importance as part of the overall quality assessment (Torjusen et al., 2001). Consumers are increasingly aware of the ecological consequences of intensive farming (Harper and Makatouni, 2002). Organic production methods and the absence of agrochemicals are thought to lower the environmental impact (Padel and Foster, 2005). Environmentally sound production has been identified as a quality property of food in a study by Wandel and Bugge (1997). As organic farmers refrain from the extensive use of fertilizers and pesticides on their fields and have to adhere to specific regulations on bio-diversity, organic produce is generally seen as being more environmentally friendly (Larceneux et al., 2012; Lee and Yun, 2015).

In case of animal products, animal welfare is believed to represent an important quality attribute too (e.g. Howard and Allen, 2006; Wier et al., 2008). Generally, consumers nowadays are more aware of those issues due to emotive media coverage and campaigns of animal welfare associations (Janet Eastwood, 1995). Ethical issues related to rearing and husbandry are multifaceted and can be linked to intensive farming practices, like the use of growth hormones to produce meat in a fast and cheap way (Janet Eastwood, 1995). Furthermore, there is a high level of support among customers for specialised animal welfare labels (Howard and Allen, 2006) which are currently discussed and introduced in different markets like Germany.

Harper and Makatouni (2002) found evidence for the significant role of animal welfare concerns in purchasing decisions of organic food. They ascertain that consumers equate organic with free-range farming and animal friendliness. Moreover, participants in their survey indicated a clear link between the happy, healthy life of livestock on the farms and the quality of their products, be it eggs or meat (Harper and Makatouni, 2002). Organic meat may also be



considered superior in quality due to the non-use of growth hormones and antibiotics (e.g. [Torjusen et al., 2001](#)). Hence, it is hypothesized that:

H 1.2: The organic label (no label) leads to higher (lower) quality evaluations in terms of environmental friendliness & animal welfare.

#### *Prestige*

Prestige was identified as an important quality dimension in a study by [Brucks et al. \(2000\)](#) referring to a product's inherent superior characteristics and its image. Goods fulfil symbolic needs, construct social identity and provide status ([Holbrook, 1999](#)). Especially in case of many credence attributes, symbolic aspects and brands become important quality characteristics as they signal a certain type of image and exclusivity ([Brucks et al., 2000](#)).

Organic has been identified before as being similar to a brand (MINTEL, 2000). Furthermore, it is considered to be a premium segment for higher socio-economic classes ([Harper and Makatouni, 2002](#); [Hill and Lynchehaun, 2002](#)). In a fMRI study, [Linder et al. \(2010\)](#) found that an organic label activated the same brain sections as luxury goods or brands. In fact, organic products have also been criticized for drawing symbolic boundaries and attracting only a wealthy, educated elite ([Johnston et al., 2011](#)). [Hill and Lynchehaun \(2002\)](#) already stated over fifteen years ago that organic food consumption is perceived fashionable as a result of media coverage, promotional activities and price premiums. It is likely that nowadays this relation is even more valid than before, considering the growing market share of ethical goods and recent food trends like veganism. Hence, organic products can be seen as status symbols ([Barrena and Sanchez, 2010](#)) and it is hypothesized that:

H 1.3: The organic label (no label) leads to higher (lower) quality evaluations in terms of prestige.

### 2.3. Value-in-Use

Having looked at how the organic label influences on customers' quality perceptions of food, this chapter turns towards the benefits that stem from the use of a product. While a product's quality attributes are generally important in buying decisions, consequences are pivotal in usage situations ([Woodruff, 1997](#)). It has been postulated that consumers derive value from quality ([Zeithaml, 1988](#)). Hence, value creation happens at the consequence level rather than the attribute one and consequences are reflected in value-in-use ([Woodruff, 1997](#)).

In the following, the concept of value and a typology of its different dimensions will be discussed. Thereafter, hypotheses will be derived on the links between the quality attributes of organic food and the respective value-in-use categories.

#### 2.3.1. Consequences as Reflected in Value-in-Use

In the means-end chain hierarchy, value is believed to be a higher level of abstraction than quality ([Zeithaml, 1988](#)).

Based on a product's quality attributes, consequences and benefits arise that stem from the consumption of those goods. They can be defined as all results originating from the use of the product, including physiological and psycho-social consequences such as satisfying hunger or enhanced self-esteem and status ([Gutman, 1982](#)). The quality characteristics and product related inferences evoked by the organic label are likely to impact on consumers' evaluation of these goods for instance in terms of their functional health benefits ([Lee and Yun, 2015](#)).

Central to the model of means-end chains is the idea, that customers choose and consume goods that produce favourable outcomes while minimizing undesired consequences. Healthy products are chosen over harmful ones because they provide more benefits and reduce risks associated with their consumption ([Gutman, 1982](#)). According to [Gutman \(1982, p. 61\)](#), "all consumer actions have consequences and consumers learn to associate particular consequences with particular actions". As consequences are generally important for the evaluation of usage situations, they are reflected in consumers' value-in-use ([Woodruff, 1997](#)).

For a long time, value was believed to be embedded in the goods that were manufactured and delivered by companies to the customer ([Grönroos, 2008](#)). This was referred to as value-in-exchange. According to [Steenkamp \(1990\)](#), however, perceived value does not stem from the purchase of a product but from its consumption, referring back to the idea that products are not cherished for their own sake but for the experience they provide ([Abbott, 1955](#)). Groceries for example are not bought to simply possess them, but rather for their ability to provide value in the form of an enjoyable, nurturing meal ([Grönroos, 2008](#)). This idea has been coined value-in-use. The concept has been defined by [Macdonald et al. \(2016, p. 98\)](#) as the "customer-perceived consequences arising from a solution that facilitate or hinder achievement of the customer's goals". This definition also includes the idea that value can never be delivered, rather it is always uniquely determined by the beneficiary in usage situations ([Vargo and Lusch, 2008](#)).

#### 2.3.2. Value Dimensions

Just like consumer goals, value-in-use proves to be a multifaceted construct (e.g. [Macdonald et al., 2016](#)). A useful typology to differentiate between the value concepts has been proposed by [Holbrook \(1994\)](#). He distinguished between 3 factors: extrinsic/intrinsic, self-oriented/other-oriented and active/reactive to create a 2x2x2 framework of consumer value. Later, he simplified the model to only include the two first mentioned factors ([Holbrook, 2006](#)), as depicted in Figure 1.

The differentiation between extrinsic and intrinsic value captures the basic idea of means-end theory. In case of extrinsic value, the consumption experience serves as a functional, instrumental means to achieve further goals, while in the case of intrinsic value the experience is appreciated for its own sake ([Holbrook, 1994](#)). Furthermore, value is considered self-oriented when it is valued for selfish reasons

|                       | <i>Extrinsic</i> | <i>Intrinsic</i> |
|-----------------------|------------------|------------------|
| <i>Self-oriented</i>  | Economic Value   | Hedonic Value    |
| <i>Other-oriented</i> | Social Value     | Altruistic Value |

**Figure 1:** Typology of Customer Value (Holbrook, 2006, p. 715)

or other-oriented when it is valued for the sake of others like family members, a community or even the planet at the macro level (Holbrook, 1994).

Based on these classifications, four distinct dimensions of value-in-use emerge that are regularly referred to in the literature (see Figure 1): economic/functional value, hedonic value, social value and altruistic value (Holbrook, 2006). However, these are not mutually exclusive as "any given consumption experience can and generally does entail many or even all of the different types of consumer value" (Holbrook, 1999, p. 186).

Other authors have found further value dimensions. Sheth et al. (1991) for example also included epistemic value that stems from the novelty of a product and conditional value which incorporates situational circumstances. In a recent contribution, Kleinaltenkamp et al. (2018) included among others environmental sustainability as a separate value-in-use dimension. In Holbrook (2006) framework this is covered by the more general category of altruistic value.

To decide on the relevant value dimensions for any analysis, one has to consider the context and type of product or service researched (Bruns and Jacob, 2016). Consumers of food products can obtain value from its taste or healthiness, but likewise from an enjoyable consumption experience or by contributing towards environmental protection (Wier et al., 2008). The consequences customers are interested in when consuming products or services are reflected in their buying motives (Grunert, 1995). Correspondingly, it can be looked at the drivers of food purchases to find out what consumers desire in terms of value.

Food consumption has been shown to be driven among others by egoistic and hedonic reasons, such as perceived benefits for one's health or pleasure and enjoyment (Step-toe et al., 1995). Moreover, altruistic considerations related to the environment and animal welfare are increasingly reflected in customers' buying motives (Lindeman and Väänänen, 2000). Last but not least, food choice is also believed to be influenced by the social image of a product (Renner et al., 2012).

Based on these findings, it was decided to refer to the four value dimensions of Holbrook (2006) framework. The quality aspects of food products as discussed in chapter 2.2 including health & safety, environmental friendliness & animal welfare, and prestige can all be related to those four value dimensions. In the following, each of them will be discussed

in detail and hypotheses on the links between the quality attributes and those value dimensions will be derived.

#### *Functional Value*

According to Sheth et al. (1991, p. 160), functional value can be defined as "the perceived utility acquired from an alternative's capacity for functional, utilitarian or physical performance [...] through the possession of salient functional, utilitarian or physical attributes". Hence, the term refers to the practical utility provided by using a specific product or service. This type of value is sought for the personal benefit, rather than a social one (Holbrook, 2006). With regards to food, this could be the intake of healthy nutritious products to maintain a healthy lifestyle (Grunert, 1995).

Health motives are often found to be the most important factor for food choices, especially in case of organic products (e.g. Padel and Foster, 2005; Zanolli and Naspetti, 2002). Health related goals can be subdivided into different categories: promotion i.e. high fiber intake (Bublitz et al., 2013), prevention i.e. pursuing a healthy lifestyle, and defence i.e. not eating harmful ingredients (Vega-Zamora et al., 2014).

The absence of chemicals and genetically modified content positively impacts on the perceived health consequences of food consumption (Baker et al., 2004; Vega-Zamora et al., 2014). Lockie et al. (2004) for example identified natural content to be related to functional aspects such as lower risks. As consumers believe natural products to be healthier, they realise short-term health benefits through their consumption (Magnusson et al., 2003). Nevertheless, some customers also link naturalness to negative associations like a worse physical appearance of fruits and vegetables (e.g. Padel and Foster, 2005). However, the functional benefits associated with the absence of additives and GM-content are likely to be more important in the context of food products.

Food safety concerns have been found to drive consumption of organic food too (e.g. Harper and Makatouni, 2002; Yiridoe et al., 2005). Generally speaking, the use of chemicals and additives is linked to unknown long-term health effects by customers (Lockie et al., 2004). Hence, avoiding those products is likely to reduce consumers' perceived risks, giving them a sense of higher security (Torjusen et al., 2001). Therefore, it is hypothesized that:

H2.1a: The higher the perceived quality of food in terms of health & safety, the higher consumers' functional value.



Food attributes linked to the environment, like the use of less pesticides and fertilizers may contribute towards the collective welfare of the community but may at the same time be beneficial for the individual, for example due to functional reasons (Harper and Makatouni, 2002; Wandel and Bugge, 1997). The non-use of pesticides or wax for instance is seen to be beneficial for one's health (Baker et al., 2004). This is especially true for fruits and vegetables as they are often eaten with their skin (Padel and Foster, 2005). In some cases, customers reported lower functional value for environmentally friendly goods e.g. for household cleaners without chemical ingredients (Green and Peloza, 2011). However, the before mentioned aspects seem to be predominant for food products.

Regarding meat products, participants in a study by Harper and Makatouni (2002) also rose concerns about the use of growth hormones or antibiotics in conventional animal rearing. Furthermore, a means-end study by Makatouni (2002) revealed that consumers see a clear correlation between animals' living conditions on the farm and their own life and health, reflected by the idea of "you are what you eat" (Makatouni, 2002, p. 351). In this sense, animal welfare provides consumers with personal i.e. functional benefits in terms of their own health (Harper and Makatouni, 2002). Hence, it is hypothesized that:

H2.2a: The higher the perceived quality of food in terms of environmental friendliness & animal welfare, the higher consumers' functional value.

#### *Hedonic Value*

Hedonic value stems from the pleasure associated with a consumption experience (Holbrook, 2006). It is derived from feelings and affections (Sheth et al., 1991). Food for example may be associated with childhood memories and could provide feelings of comfort (Sheth et al., 1991).

Besides being related to functional value, health & safety aspects of food products are also likely to be linked to hedonic value, as food additives and chemicals impact on how customers feel about a product (Lee and Yun, 2015; Padel and Foster, 2005). In a study by Baker et al. (2004), the avoidance of unnatural content was found to be related to better taste and higher state of enjoyment when consuming. Lockie et al. (2004) identified natural content to evoke emotional appeal and pleasure among consumers. Moreover, in a study by Vega-Zamora et al. (2014), naturalness was associated with pleasant feelings and evocations of childhood and nature. Hence,

H2.1b The higher the perceived quality of food in terms of health & safety, the higher consumers' hedonic value.

Pickett-Baker and Ozaki (2008) found that people felt good about using products that were less harmful for the environment. Supporting the idea of a hedonic satisfaction, De Young (1996) posited, that people engage in sustainable behaviours for their personal contentment. It has been

shown that in case of organic food, consumers may get emotional satisfaction through its consumption as they feel to contribute to the well-being of animals and the wider environment (Harper and Makatouni, 2002, p. 297). But even for conventional products, animal welfare remains a highly emotional issue and is thus likely to impact on the emotional benefits associated with its consumption (Janet Eastwood, 1995). Accordingly, it is hypothesized that:

H2.2b: The higher the perceived quality of food in terms of environmental friendliness & animal welfare, the higher consumers' hedonic value.

#### *Altruistic Value*

Altruistic value occurs from the pursuit of doing something for the sake of others rather than for one's own benefit. It encompasses ethical as well as spiritual issues and is characterized as other-oriented in the framework of Holbrook (2006). With regards to food products, the belief to contribute to environmental protection or to support better housing conditions for animals may provide customers with this type of value.

When buying environmentally friendly goods consumers can get positive rewarding feelings from that activity (Pickett-Baker and Ozaki, 2008). It has been argued before that consuming altruistically in line with one's moral principles provides customers with a "warm glow" (Andreoni, 1990, p. 464). This might be especially important for buyers of organic products who feel like they contribute to something better and act morally right (Dean et al., 2008). In fact, means-end studies on organic food consumption regularly identify links between the use of organics and consumers universal values of benevolence and altruism (e.g. Makatouni, 2002; Zanolli and Naspetti, 2002).

Generally speaking, customers have become increasingly aware of the inhumane treatment of animals in industrialized farming due to emotive media coverage (Janet Eastwood, 1995). Hence, especially in case of animal products, altruistic benefits like the support of better animal housing conditions are sought for by customers. In a study by Baker et al. (2004) for example, taking responsibility for other creatures was identified as an important higher-order goal that was valued by consumers. Therefore, it is hypothesized that:

H2.2c: The higher the perceived quality of food in terms of environmental friendliness & animal welfare, the higher consumers' altruistic value.

#### *Social Value*

Social value stems from customers' associations of a good with specific reference groups (Sheth et al., 1991). It has been defined as "the utility derived from a product's ability to enhance social self-concept" (Sweeney and Soutar, 2001, p. 211) and is related to self-image, prestige and status (Holbrook, 1994). Feeling accepted and appreciated can constitute a source of value (Grönroos, 2008) especially in the presence of strong normative beliefs that have been defined by

Ajzen (1991, S. 195) as an individual's perception of the approval of a behaviour by a so-called "important others". Consequently, when the use of specific products enhances one's status, social value occurs (Holbrook, 2006).

In means-end theory, psychological and sociological consequences such as self-esteem or status are considered important aspects linking products to end-goals (Gutman, 1982). There appears to be a concurrence between a person's lifestyle and the goods he or she purchases. Products have long been regarded as carrying personal and social meaning (Levy, 1959). From that perspective, they are not seen as objects but rather as symbols used for distinction and proof of social positions.

It has been argued before that buying ethical products or adopting a certain diet like veganism may also be driven by the desire to be associated with a particular group (Bublitz et al., 2010). This might be especially true for organic produce, that is associated with higher prices and prestige (Harper and Makatouni, 2002). It is considered a status symbol that may lead to enhancement in the eyes of others (Barrena and Sanchez, 2010; Green and Peloza, 2011). In fact, Johnston et al. (2011) showed that privileged groups constructed symbolic boundaries around ethical food consumption to show themselves as behaving morally right. Eating high-quality products was seen by participants as providing status and cultural distinction. Hence, it is hypothesized that:

H2.3: The higher the perceived quality of food in terms of prestige, the higher consumers' social value.

## 2.4. Well-Being

While the previous chapter looked at the links between specific quality dimensions of food products and value-in-use perceptions, this chapter presents the concept of well-being as an end that customers try to achieve through the consumption of high quality goods (e.g. Padel and Foster, 2005; Zanolli and Naspetti, 2002). First, the concept will be explained in general with a specific focus on the work environment. Then, the rather new research stream of food well-being will be presented and used in order to derive hypotheses on the links between value-in-use and well-being with regards to food consumption.

### 2.4.1. Well-Being as an End-Goal and Societal Value

It is important to note the difference between the concepts of value and values in the plural form as both are distinct yet related to one another. Value refers to the outcome of a consumer's evaluative judgement of the consequences associated with a consumption situation as described before, whereas values are the underlying norms, ideals or goals that form the basis of the underlying evaluation criteria (Holbrook, 1999). In the words of Gutman (1982, p. 61): "values give consequences valence and importance".

Values can be defined as specific end-states of existence that are preferred over others. They stem from culture, social surroundings and personality and are shaped by institutions existing in any society (Rokeach, 1973). Values serve as guiding principles in everyday life and may vary in their relative importance depending on the situation. In studies on organic food consumption well-being is commonly identified as an important end-goal (e.g. Padel and Foster, 2005; Zanolli and Naspetti, 2002).

Even though a common definition of the term well-being is missing so far, it is agreed that it is a broad concept that includes much more than just physical health (Diener, 2006). According to the Dictionary (1989, p.133), well-being is "the state of being or doing well in life; happy, healthy, or prosperous condition; moral or physical welfare (of a person or community)". What is important though, is that the term is subjective in its nature as it is an umbrella term for an individual's evaluation regarding his or her life, body and mind (Diener, 2006). The evaluation can either occur at a given moment or over a longer period of time and normally includes cognitive and affective components, ranging from work satisfaction to happiness and fulfilment (Diener, 2006). Marks and Shah (2004) further denote personal development and contribution to society as important constituents of the construct. Taken together, subjective well-being has been referred to as a measure for both, individual and societal quality of life (Diener, 2006).

The concept is of increasing interest for various stakeholders, including managers, politicians and supranational organisations due to the benefits associated with it (Diener, 2006). An individual's state of well-being affects the cognitive and emotional functioning of the being and may also have positive spill-over effects on other people. Indicators of subjective well-being provide useful insights that may inform policies for work life, social services, recreation and the environment (Diener, 2006).

Besides in global measures of life satisfaction, well-being indicators can further be used in specific domains such as the work environment (Diener, 2006). HR-managers in many companies are committed to foster the topic by regularly gathering and evaluating data on well-being through surveys (Grant et al., 2007). Their focus on this topic seems to be justified by its significant impact on job performance and motivation as well as absenteeism and turnover rates (Tenney et al., 2016).

Well-being related interventions may be directed towards the individual or the collective and can be classified proactive, e.g. fitness classes or reactive measures in case of massages (Evans and Prilleltensky, 2007). Important in this regard is also the provision of healthy, nutritious meals that satisfies employees' needs and gives them hedonic pleasure. Access to nutritious food and a clean environment were identified before as signs of collective well-being, that are beneficial for the individual too (Evans and Prilleltensky, 2007). In fact, it has been proposed that workplace lunches contribute to employee well-being (e.g. Haugaard et al., 2016; Price et al., 2016). Hence, in the following, it will be looked at

how food influences on consumer well-being.

#### 2.4.2. Food Well-Being

Food well-being (FWB) is a rather recent stream of transformative consumer research which focuses on analysing the impacts of food choices on health and subjective well-being (Bublitz et al., 2013). There has been a growing interest into that topic in recent years due to the mounting number of so-called civilization diseases (Ares et al., 2014). The concept of FWB is linked to the diverse consequences people experience when consuming food, ranging from physiological to psychological and social ones (Block et al., 2011). It refers to the post-prandial state after food consumption and includes dimensions such as satisfaction, relaxation, energy or sleepiness (Boelsma et al., 2010).

Indeed, well-being has been identified as an important motive and end-goal in the means-end hierarchy of organic food consumption (e.g. Padel and Foster, 2005; Zanolini and Naspetti, 2002). Prior research has shown that labelling which refers to production methods, environmental impact or animal welfare may enhance consumer well-being (Block et al., 2011). Moreover, Yiridoe et al. (2005) reasoned that the quality attributes of organic products can be considered input factors of a consumer's demand function for enhanced well-being. The positive link between organic food and well-being may either stem from the physiological reasoning of organics being healthier or from a purely cognitive perception based on consumers' beliefs and expectations of labelled products (Apaolaza et al., 2018). In the latter case, the influence of food consumption on subjective well-being may be attributed solely to the halo effect of the organic label. In fact, in their study, Apaolaza et al. (2018) also tested this effect empirically and found highly significant effects on the proposed causal relationships between the consumption of organically labelled food and subjective well-being.

Generally speaking, it is believed that food choices strongly affect subjective well-being (Ares et al., 2014). In their study, the authors showed that the effects of food consumption on well-being were related to hedonic aspects, such as pleasure but also physical health and personal fulfilment. When participants were prompted to think about top of mind associations on food and well-being, they mostly mentioned health as well as fruits and vegetables (Ares et al., 2014). Important food characteristics were vitamins, fibre intake and naturalness with food additives being negatively linked to perceived well-being. Food choice was thought to be related to intellectual dimensions such as participants' mental performance, concentration and sleepiness but likewise to psychological ones like pleasure, good mood, energy and vitality (Ares et al., 2014).

Even if the results of Ares et al. (2014) were only derived from a group of 120 individuals with similar sociocultural backgrounds, participants' associations and conceptualisation of food related well-being provide valuable insights. When consumers perceive meals to be healthy, nutritious and without additives, its consumption could have a positive impact on their subjective well-being (Apaolaza et al., 2018).

However, in this research it is argued that it is not the quality attributes of food products that directly impact on well-being but rather, in accordance with means-end chain theory, the consequences that arise from these attributes in usage situations. Bublitz et al. (2013) for example contended that FWB stems from functional, hedonic and symbolic consequences. Grant et al. (2007) further referred to eudaemonic fulfilment and social aspects as being important in workplace settings. Thus, subjective post-prandial well-being is likely to be influenced by all of the value dimensions of food products. Hence it is hypothesized that:

H3.1: The higher the functional value-in-use of food, the higher consumers' subjective well-being in the post-prandial state.

H3.2: The higher the hedonic value-in-use of food, the higher consumers' subjective well-being in the post-prandial state.

H3.3: The higher the altruistic value-in-use of food, the higher consumers' subjective well-being in the post-prandial state.

H3.4: The higher the social value-in-use of food, the higher consumers' subjective well-being in the post-prandial state.

#### 2.5. Research Model

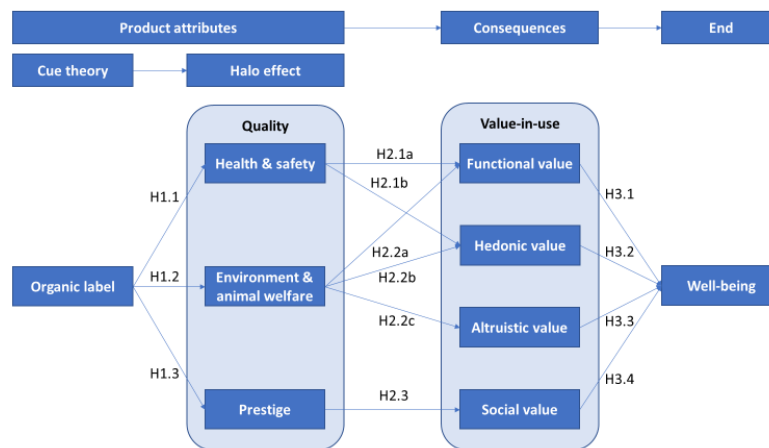
The hypothesized relationships between the latent variables are summarized in the research model (Figure 2). The organic label as an extrinsic cue is believed to positively impact on consumers' quality evaluations in terms of health & safety, environment & animal welfare as well as prestige through the halo effect. These quality perceptions are likely to lead to specific consequences, reflected in the four value-in-use dimensions by Holbrook (2006). Finally, these value categories are hypothesized to positively contribute to consumers' subjective state of well-being after consumption.

### 3. Methodology

Having discussed the literature on organic food consumption, value-in-use and well-being, this chapter describes the methodology that was chosen in order to answer the research questions and to prove the hypothesized links between the constructs in the conceptual model. First, the study design, the pre-test and the data collection methods will be explained. Thereafter, the sample characteristics and the items used to measure the constructs will be presented. The chapter closes with a short description on the method of analysis.

#### 3.1. Study Design

In this research project, a quantitative study design was applied that allowed for an efficient collection of data from many participants and a convenient statistical analysis (Saunders et al., 2009). Throughout the research process,



**Figure 2:** Research Model

the collected data was treated confidentially. It was only used for the purpose of this project and will be deleted upon completion. Before starting the survey, every participant received information on the topic of the study and was informed about data privacy. The participation was voluntary and every participant was able to withdraw from answering the questions at any time. However, in order to ensure a high response rate, it was decided to incentivise participation by raffling three Amazon vouchers among all participants who voluntarily consigned their e-mail addresses.

The survey was conducted online in a virtual canteen setting. At the beginning of the questionnaire, respondents were randomly allocated to two groups. Both groups received a short description of the canteen setting. Participants were asked to project themselves into the situation before answering the questions. Subsequently, three exemplary meals were shown, which were adapted from the menu charts of *Apetito* (2018), one of the biggest food service providers in Europe. Two meals with meat were included: pork escalope with spaetzle and mixed vegetables as well as chicken breast with rice, cauliflower and carrots. Moreover, one vegetarian option was added: spinach cappelletti with pesto sauce.

Both groups received the same meals and pictures, however, the experimental group was treated with a verbal description and visual stimuli of the German national organic label in combination with the European logo (see Figure 3). This single-cue design secured that all other stimuli were held constant so that differences between the two groups could be traced back to the label. Organic standards were not explained in the survey, as participants' beliefs about the meaning of the label and its impact on quality and value perceptions were of interest. A manipulation check was placed at the end of the questionnaire in order not to prime the answers of both groups. The full survey can be found in the Appendix.

### 3.2. Pre-Test

Before running the survey, the questionnaire was pre-tested between the 24th of May and the 1st of June 2018 by

twelve respondents, focusing on comprehensibility and orthography. Additionally, cognitive interviewing was applied with two participants. They were asked to fill in the survey in the presence of the researcher and to voice all concerns or thoughts they had while answering the questions. This procedure ensured, that all items were interpreted and understood as intended (Bruns and Jacob, 2016). The feedback of these pre-tests led to slight adjustments in the wording of some items. Furthermore, one of the meals was renamed as both participants of the cognitive interviews were unsure about the meaning of the word cappelletti, hence this was changed to the more well-known term ravioli.

### 3.3. Data Collection and Sample

For this research project, the data was primarily collected online through a self-administered web survey using Qualtrics. A print-out version was also derived to reach more participants. However, the focus of the data collection was on the online survey. This was considered to be most convenient for the respondents who could answer the questions wherever and whenever they wanted. One drawback, that is commonly referred to when gathering data through an online questionnaire is that the answers are standardised which does not allow for in-depth understanding (Saunders et al., 2009). However, given the time-constraints of this research, this data collection method was considered to be most effective in order to reach a large sample size.

The data was gathered between the 5th of June and the 1st of July 2018. The survey was distributed via mail and through social media channels but handed out in print at different occasions, too. Hence, a mix of self-selection and convenience sampling was applied in this research. This ensured high response rates but led to an unrepresentative sample so that the findings cannot easily be generalised (Saunders et al., 2009). Nevertheless, this sampling strategy was chosen to get as many people as possible to answer the questionnaire thereby also enhancing the representativeness of the findings.



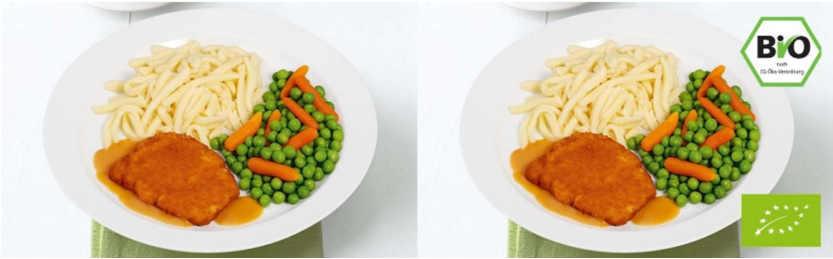


Figure 3: Manipulation of the Food Offer (Pictures adapted from [Apetito, 2018](#))

Table 1: Sample Characteristics

|              |        | Total sample |            | Experimental group |            | Control group |            |
|--------------|--------|--------------|------------|--------------------|------------|---------------|------------|
| Demographics |        | N            | Rel. Freq. | N                  | Rel. Freq. | N             | Rel. Freq. |
| Gender       | Female | 104          | 61%        | 55                 | 62%        | 49            | 60%        |
|              | Male   | 66           | 39%        | 33                 | 37%        | 33            | 40%        |
|              | N/A    | 1            | -          | 1                  | 1%         | -             | -          |
|              | Total  | 171          | 100%       | 89                 | 100%       | 82            | 100%       |
| Age group    | <20    | 7            | 4%         | 3                  | 3%         | 4             | 5%         |
|              | 20-30  | 97           | 57%        | 53                 | 60%        | 44            | 54%        |
|              | 31-45  | 20           | 12%        | 10                 | 11%        | 10            | 12%        |
|              | 46-60  | 35           | 20%        | 17                 | 19%        | 18            | 22%        |
|              | >60    | 8            | 5%         | 4                  | 5%         | 4             | 5%         |
|              | N/A    | 4            | 2%         | 2                  | 2%         | 2             | 2%         |
|              | Total  | 171          | 100%       | 89                 | 100%       | 82            | 100%       |

Altogether, 213 responses could be collected that were closely examined and cleansed before the subsequent analysis. In order to deal with missing value problems, a forced-answer approach was used to make participants answer all questions, risking that some respondents might then exit the survey ([Hair Jr. et al., 2014](#)). In fact, 33 responses had missing values above 15%. These were excluded from the data set. Moreover, in nine cases, the manipulation was not successful. The manipulation was considered unsuccessful when respondents in the control group agreed or strongly agreed to the statement that organic food was offered in this canteen, and vice versa for the experimental group. Last but not least, one answer was excluded due to implausible answer patterns.

Hence, taken together, 171 valid responses could be collected. The total sample consist of 61% female and 39% male participants ranging from 14 to 93 years with an average age of 34,45. The demographic profiles of the two groups are rather similar as can be seen from the data in [Table 1](#).

3.4. Measurement Items

In order to assess the latent variables in the model, it was referred to established scales on food quality dimensions, value-in-use and well-being. As all constructs were measured reflectively, the items were interchangeable and highly correlated so that items could be left out without changing the meaning of the construct ([Hair Jr. et al., 2014](#)).

Participants’ quality perceptions of the meals offered in the canteen were recorded by using items referring to their

beliefs about the foods’ attributes (see [Table 2](#)). Health & safety qualities were measured using five items from [Step-toe et al. \(1995\)](#), three items of [Lockie et al. \(2002\)](#) and two items by [Renner et al. \(2012\)](#). Those also included specific items on naturalness, which, as argued in [Chapter 2.2.4.1](#), was considered a subcategory of health & safety in this survey. The items on environmental quality & animal welfare were taken from [Renner et al. \(2012\)](#) and [Lindeman and Väänänen \(2000\)](#), while prestige was measured with three items of [Renner et al. \(2012\)](#) and one item from [Sweeney and Soutar \(2001\)](#).

Value-in-use scales were adapted from [Sweeney and Soutar \(2001\)](#) for hedonic and social value and [Magnusson et al. \(2003\)](#) for functional value. Here, participants were asked to indicate their agreement with statements on the benefits associated with the consumption of the canteen food. The same applied for altruistic value which was assessed with three items by [Dean et al. \(2008\)](#) and two items from [Kleinaltenkamp et al. \(2018\)](#).

In the end, respondents were requested to answer six questions on their perceived post-prandial well-being, which were taken from [Boelsma et al. \(2010\)](#). All items were measured on a symmetric, equidistant 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). The wording of all items was slightly adjusted to fit to the context of this study. All answers were coded numerically to allow for an application of multivariate analysis. A detailed code plan can be found in the [Appendix](#).



**Table 2:** Measurement Items

| Construct                              | Items                                                                    | Adapted from                                         |
|----------------------------------------|--------------------------------------------------------------------------|------------------------------------------------------|
| Health & safety                        | ... contains a lot of vitamins and minerals                              | Steptoe et al. (1995)                                |
|                                        | ... is nutritious                                                        |                                                      |
|                                        | ... is low in calories                                                   |                                                      |
|                                        | ... contains no additives/artificial ingredients                         | Lockie et al. (2002)                                 |
|                                        | ... contains only natural ingredients                                    |                                                      |
|                                        | ... is healthy to eat                                                    |                                                      |
|                                        | ... has low chemical residues (e.g. pesticides)                          | Renner et al. (2012)                                 |
|                                        | ... is prepared in a way that preserves its natural goodness             |                                                      |
|                                        | ... is safe to consume                                                   |                                                      |
| Environmental quality & animal welfare | ... is natural (e.g. not genetically modified)                           | Renner et al. (2012)<br>Lindeman and Väänänen (2000) |
|                                        | ... is environmentally friendly (e.g. production, transport)             |                                                      |
|                                        | ... have been prepared in an environmentally friendly way                |                                                      |
|                                        | ... have been produced in a way that animals have not experienced pain   |                                                      |
| Prestige                               | ... have been produced in a way that animals' rights have been respected | Renner et al. (2012)                                 |
|                                        | ... is trendy                                                            |                                                      |
|                                        | ... is liked by others                                                   |                                                      |
|                                        | ... can be considered to be special                                      |                                                      |
| Functional value                       | ... has a positive social image                                          | Sweeney and Soutar (2001)                            |
|                                        | ... I improve my health                                                  |                                                      |
|                                        | ... gives me a good conscience                                           |                                                      |
|                                        | ... I avoid food related risks                                           |                                                      |
| Hedonic value                          | ... reduces the risk for illness                                         | Magnusson et al. (2003)                              |
|                                        | ... I enjoy it                                                           |                                                      |
|                                        | ... I feel relaxed about it                                              |                                                      |
|                                        | ... it makes me feel good                                                |                                                      |
| Social value                           | ... it gives me pleasure                                                 | Sweeney and Soutar (2001)                            |
|                                        | ... improves the way I am perceived                                      |                                                      |
|                                        | ... would help me feel accepted                                          |                                                      |
|                                        | ... would make a good impression on other people                         |                                                      |
| Altruistic value                       | ... would give me social approval                                        | Sweeney and Soutar (2001)                            |
|                                        | ... I contribute to something better                                     |                                                      |
|                                        | ... it feels like the morally right thing                                |                                                      |
|                                        | ... I feel like a better person                                          |                                                      |
| Well-being                             | ... I contribute to environmental protection                             | Dean et al. (2008)                                   |
|                                        | ... I reduce environmental pollution                                     |                                                      |
|                                        | ... I feel pleasant                                                      | Boelsma et al. (2010)                                |
| Well-being                             | ... I feel satisfied                                                     |                                                      |
|                                        | ... I feel relaxed                                                       |                                                      |
|                                        | ... I feel sleepy                                                        |                                                      |
|                                        | ... I feel physically energetic                                          |                                                      |
|                                        | ... I feel mentally alert                                                |                                                      |

### 3.5. Method of Analysis

In order to analyse significant differences between the experimental and the control group in terms of their quality perceptions, a two-tailed t-test was used. For that purpose, participants' evaluation of quality in terms of health & safety, environmental quality, animal welfare and prestige

were compared between the groups. This allowed to substantiate the hypothesized halo effect of the organic label on consumers' quality perception.

Afterwards, partial least squares structural equation modelling (PLS-SEM) was applied to assess the relationships between the latent variables in the model. SEM is an advanced

statistical analysis method that combines techniques of regression and factor analysis allowing for the examination of multiple relationships simultaneously (Hair Jr. et al., 2014). The PLS approach is the appropriate method for testing complex models and can be applied even in case of small sample sizes (Hair Jr. et al., 2014). Moreover, it is highly robust to missing values and does not require a specific distribution of the data. Thus, it can be applied in many research settings especially for testing many structural relationships at the same time (Hair Jr. et al., 2014). For the purpose of analysis, SmartPLS version 3.2.7 was used as a statistical software.

## 4. Results

Having described the methodology applied in this research, this chapter now turns towards the findings of the survey. First up, the results of the organic label halo effect on consumers' quality evaluations will be presented. Thereafter, the results of the SEM analysis will be outlined.

### 4.1. Results of the Organic Label Halo Effect

An overview on the descriptive statistics of the overall sample can be found in Table 3. Correlations between the constructs were calculated and can be found therein.

The results of the two-tailed t-test are summarized in Table 4. The assumptions of variance analysis like homogeneity of variances and normal distribution have been tested and verified. The treatment of the experimental group was successful, as indicated by the significant difference between the two groups with regards to the manipulation check question. While the experimental group on average agreed with the statement that organic food was offered in the canteen ( $\mu=5.71$ ), the control group did not ( $\mu=3.31$ ).

In order to test for group differences between the quality dimensions, an average of the respective items for each construct was calculated. The group treated with the organic label evaluated the food significantly higher ( $\mu=4.56$ ) in terms of health & safety than the control group ( $\mu=3.48$ ). Therefore, H1.1 is supported by the data. Similarly, the results for environmental quality & animal welfare as well as prestige show significant differences between the groups at  $\alpha$ -levels of at least 5%. In both cases, respondents in the organic setting evaluated the food offer significantly higher in comparison to the ones in the normal setting. Hence, H1.2 and H1.3 can be confirmed.

Table 4 further includes data for the evaluations of value-in-use and well-being by both groups. All differences were found to be highly significant. To test the hypothesized relationships between the constructs subsequently PLS-SEM was applied.

### 4.2. Results of the SEM Modelling

The analysis of the SEM results is two-folded. First, the measurement models have to be tested for reliability and validity. Then, in a second step, the structural model can be as-

sessed and tested for the hypothesized relationships between the constructs (Hair Jr. et al., 2014).

#### 4.2.1. Assessment of the Measurement Model

As described before, in this research all constructs were measured reflectively. In order to evaluate the reflective measurement models, composite reliability as well as convergent and discriminant validity were considered. First, the outer loadings between the latent variables and their reflective items were derived from the analysis to test for the reliability of the measures. The results can be found in Table 5. Three items with outer loadings of less than 0.7 were excluded from the analysis following Hair Jr. et al. (2014).

To assess internal consistency, composite reliability (CR) was considered. The recommended threshold for CR is at least 0.6 or 0.7 in more advanced stages of research (Hair Jr. et al., 2014). In this study, all CR values reach satisfactory levels above 0.85 (see Table 6).

Convergent validity was assessed via average variance extracted (AVE). The threshold for AVE was set to 0.5 as commonly suggested, to secure that the constructs explain at least half of the variance of their indicators (Chin, 2010). The results show that AVE for all constructs is above this threshold.

Last but not least, the Fornell-Larcker criterion was used to test for discriminant validity. The results in Table 7 show no correlation between any two constructs higher than the square root of the respective AVE. Hence, discriminant validity is supported by the data and the analysis can move on with assessing the structural model.

#### 4.2.2. Assessment of the Structural Model

Having secured the reliability and validity of the construct measures, the theoretical model can be examined. To assess the predictive power of the model, it was looked at the  $R^2$ -values of the endogenous variables in the model in order to see how much of their variance is explained by the hypothesized relationships in the model. PLS-SEM tries to maximize  $R^2$ -values of the endogenous latent variables (Chin, 2010). In consumer research, values above 0.2 can be considered high (Hair Jr. et al., 2014).

The model is able to explain over 50% of the variance for functional ( $R^2=0.58$ ) and altruistic value ( $R^2=0.538$ ). Moreover, 45.5% of the variance in hedonic value and 19.3% for social value respectively can be explained by the data. Ultimately the model is able to explain 60.3% of the variance in well-being. Hence, the  $R^2$ -scores can be considered satisfactory (see Table 9).

Next, the path coefficients were analysed to assess the hypothesized effects between the latent variables. However, before being able to test for the significance and relevance of the relationships, collinearity needed to be looked at (Hair Jr. et al., 2014). All VIF values are below the threshold of 5, consequently, collinearity was not indicated and the bootstrapping routine could be applied. The number of bootstrap samples was set to 1,000. The path coefficients and their respective significance levels are presented in Table ??.

**Table 3:** Descriptive Statistics and Correlation

| Construct                        | Mean | SD   | Correlation |      |      |      |      |      |      |     |
|----------------------------------|------|------|-------------|------|------|------|------|------|------|-----|
|                                  |      |      | (1)         | (2)  | (3)  | (4)  | (5)  | (6)  | (7)  | (8) |
| Health & safety (1)              | 4.03 | 1.19 | 1           |      |      |      |      |      |      |     |
| Environment & animal welfare (2) | 3.62 | 1.36 | 0.79        | 1    |      |      |      |      |      |     |
| Prestige (3)                     | 4.05 | 1.03 | 0.53        | 0.44 | 1    |      |      |      |      |     |
| Functional value (4)             | 3.08 | 1.03 | 0.72        | 0.70 | 0.46 | 1    |      |      |      |     |
| Hedonic value (5)                | 4.31 | 1.16 | 0.66        | 0.61 | 0.47 | 0.67 | 1    |      |      |     |
| Altruistic value (6)             | 3.54 | 1.31 | 0.67        | 0.72 | 0.44 | 0.83 | 0.65 | 1    |      |     |
| Social value (7)                 | 3.79 | 1.17 | 0.51        | 0.44 | 0.43 | 0.69 | 0.55 | 0.67 | 1    |     |
| Well-being (8)                   | 4.14 | 0.85 | 0.62        | 0.54 | 0.47 | 0.67 | 0.69 | 0.61 | 0.60 | 1   |

**Table 4:** Results of the t-Test for Group Differences

Note: \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$

| Construct                         | Organic group | Control group | t-test    |         | Manipulation/ Hypothesis |
|-----------------------------------|---------------|---------------|-----------|---------|--------------------------|
|                                   |               |               | t-Wert    | p-Wert  |                          |
| Manipulation check: Organic food? | 5.71          | 3.31          | -12.14*** | <0.0001 | Manipulation successful  |
| Health & safety                   | 4.56          | 3.48          | -6.66***  | <0.0001 | H1.1 supported           |
| Environment & animal welfare      | 4.31          | 2.89          | -7.87***  | <0.0001 | H1.2 supported           |
| Prestige                          | 4.23          | 3.86          | -2.34**   | 0.02    | H1.3 supported           |
| Functional value                  | 3.54          | 2.59          | -6.77***  | <0.0001 |                          |
| Hedonic value                     | 4.61          | 3.99          | -3.68***  | 0.0003  |                          |
| Altruistic value                  | 4.13          | 2.92          | -6.76***  | <0.0001 |                          |
| Social value                      | 4.15          | 3.41          | -4.31***  | <0.0001 |                          |
| Well-being                        | 4.35          | 3.91          | -3.52***  | 0.0006  |                          |

As expected, the results show a significant positive effect of health & safety on functional and hedonic value, thus providing support for H2.1a+b. Moreover, as hypothesized in H2.2a-c, the quality of food products in terms of environment & animal welfare positively contributes to functional, hedonic and altruistic value. All of those paths are highly significant at  $\alpha$ -levels of less than 1%. Additionally, prestige is found to significantly impact on social value, thereby supporting hypothesis H2.3.

Last but not least, well-being is found to be driven by hedonic and functional value, however, the hypothesized links to altruistic and social value are insignificant, thus, H3.3 and H3.4 cannot be supported. The final model, including all path coefficients can be found in Figure 4. Dotted lines indicate insignificant relationships.

## 5. Discussion

In this section, the findings as presented in the last section will be critically evaluated in reference to existing literature. The chapter is structured according to the research questions and will thus start with an examination of the organic label halo effect. It will then go on to discuss the links between quality and value for food products and ultimately end with analysing the impact on post-prandial well-being.

### 5.1. The Organic Label Halo Effect in a Canteen Setting

Several researchers have identified organic products to be perceived higher in quality in comparison to conventional alternatives (e.g. Hill and Lynchehaun, 2002; Larceneux et al., 2012). This inference has been attributed to the so-called halo effect (Thorndike, 1920). This survey aimed at analysing the organic label halo effect on consumers' quality perceptions in a canteen setting. The results show support for its existence.

The food in the experimental setting is considered superior in terms of healthiness and nutritional value, supporting previous findings by Magnusson et al. (2001) and Hoefkens et al. (2009). Moreover, through inferences about natural production methods and the use of less artificial ingredients, participants in this study believe the labelled meals to be safer. Due to the prohibition of artificial pesticides and fertilizers in farming, organic food is certain to contain less residues and nitrate levels, particularly in fruits and vegetables (Hoefkens et al., 2009). However, it can be argued that all food products are strictly regulated to ensure they are safe to consume. Besides, only inconclusive evidence exists, that organic produce is healthier or more nutritious in terms of vitamins and minerals than conventional alternatives (Yiridoe et al., 2005). Hence, consumers' perception of organic food being healthier can be attributed to inferential beliefs evoked by the label. Especially in case of many credence attributes, the label might in fact be the only available cue for consumers' quality evaluations (Caswell and Mojduszka,

**Table 5:** Outer Model Loadings

\* Items were excluded from further analysis

| Items     | Constructs      |                   |          |                  |               |                  | Social value | Well-being |
|-----------|-----------------|-------------------|----------|------------------|---------------|------------------|--------------|------------|
|           | Health & safety | Environ.& animals | Prestige | Functional value | Hedonic value | Altruistic value |              |            |
| Health_1  | 0.814           |                   |          |                  |               |                  |              |            |
| Health_2  | 0.708           |                   |          |                  |               |                  |              |            |
| Health_3  | 0.723           |                   |          |                  |               |                  |              |            |
| Health_4  | 0.841           |                   |          |                  |               |                  |              |            |
| Health_5  | 0.610*          |                   |          |                  |               |                  |              |            |
| Health_6  | 0.746           |                   |          |                  |               |                  |              |            |
| Natural_1 | 0.851           |                   |          |                  |               |                  |              |            |
| Natural_2 | 0.870           |                   |          |                  |               |                  |              |            |
| Natural_3 | 0.787           |                   |          |                  |               |                  |              |            |
| Natural_4 | 0.843           |                   |          |                  |               |                  |              |            |
| Animal_1  |                 | 0.860             |          |                  |               |                  |              |            |
| Animal_2  |                 | 0.899             |          |                  |               |                  |              |            |
| Environ_1 |                 | 0.850             |          |                  |               |                  |              |            |
| Environ_2 |                 | 0.877             |          |                  |               |                  |              |            |
| Prest_1   |                 |                   | 0.822    |                  |               |                  |              |            |
| Prest_2   |                 |                   | 0.723    |                  |               |                  |              |            |
| Prest_3   |                 |                   | 0.762    |                  |               |                  |              |            |
| Prest_4   |                 |                   | 0.752    |                  |               |                  |              |            |
| Funct_1   |                 |                   |          | 0.909            |               |                  |              |            |
| Funct_2   |                 |                   |          | 0.826            |               |                  |              |            |
| Funct_3   |                 |                   |          | 0.845            |               |                  |              |            |
| Funct_4   |                 |                   |          | 0.876            |               |                  |              |            |
| Hedon_1   |                 |                   |          |                  | 0.850         |                  |              |            |
| Hedon_2   |                 |                   |          |                  | 0.876         |                  |              |            |
| Hedon_3   |                 |                   |          |                  | 0.853         |                  |              |            |
| Hedon_4   |                 |                   |          |                  | 0.888         |                  |              |            |
| Altru_1   |                 |                   |          |                  |               | 0.802            |              |            |
| Altru_2   |                 |                   |          |                  |               | 0.911            |              |            |
| Altru_3   |                 |                   |          |                  |               | 0.817            |              |            |
| Altru_4   |                 |                   |          |                  |               | 0.883            |              |            |
| Altru_5   |                 |                   |          |                  |               | 0.852            |              |            |
| Social_1  |                 |                   |          |                  |               |                  | 0.866        |            |
| Social_2  |                 |                   |          |                  |               |                  | 0.662*       |            |
| Social_3  |                 |                   |          |                  |               |                  | 0.919        |            |
| Social_4  |                 |                   |          |                  |               |                  | 0.862        |            |
| Well_1    |                 |                   |          |                  |               |                  |              | 0.897      |
| Well_2    |                 |                   |          |                  |               |                  |              | 0.806      |
| Well_3    |                 |                   |          |                  |               |                  |              | 0.768      |
| Well_4    |                 |                   |          |                  |               |                  |              | 0.460*     |
| Well_5    |                 |                   |          |                  |               |                  |              | 0.846      |
| Well_6    |                 |                   |          |                  |               |                  |              | 0.764      |

1996). It can be summarized that the organic label positively biases their quality evaluations in terms of health & safety in a canteen setting, thereby supporting findings of previous research (e.g. Apaolaza et al., 2017; Lee et al., 2013).

Similar results could be found for environmental friendliness & animal welfare that are believed to form important quality aspects of food products (Torjusen et al., 2001). Even though it has been claimed that customers are often un-

aware of the exact standards of ecological production methods (e.g. Hill and Lynchehaun, 2002; Yiridoe et al., 2005), they equate organic with environmental and animal friendliness (Harper and Makatouni, 2002). Therefore, those products are generally perceived to be better for the environment (e.g. Larceneux et al., 2012; Lee and Yun, 2015). This could also be shown by the results of this survey. The food offered in the organic setting is evaluated significantly higher

**Table 6:** Assessment of the Measurement Model

(1) Cronbach's alpha (2) Composite reliability (c) Average variance extracted

| Construct                    | No. items | CRA (1) (>0,7) | CR (2) (>0,7) | AVE (3) >0,5) |
|------------------------------|-----------|----------------|---------------|---------------|
| Health & safety              | 9         | 0.930          | 0.942         | 0.645         |
| Environment & animal welfare | 4         | 0.895          | 0.927         | 0.760         |
| Prestige                     | 4         | 0.764          | 0.850         | 0.586         |
| Functional value             | 4         | 0.887          | 0.922         | 0.748         |
| Hedonic value                | 4         | 0.890          | 0.924         | 0.751         |
| Altruistic value             | 4         | 0.907          | 0.931         | 0.729         |
| Social value                 | 3         | 0.884          | 0.928         | 0.811         |
| Well-being                   | 5         | 0.878          | 0.912         | 0.674         |

**Table 7:** Fornell-Larcker Criterion

| Construct                        | (1)   | (2)   | (3)   | (4)   | (5)   | (6)   | (7)   | (8)   |
|----------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Health & safety (1)              | 0.803 |       |       |       |       |       |       |       |
| Environment & animal welfare (2) | 0.786 | 0.872 |       |       |       |       |       |       |
| Prestige (3)                     | 0.545 | 0.438 | 0.766 |       |       |       |       |       |
| Functional value (4)             | 0.726 | 0.713 | 0.460 | 0.865 |       |       |       |       |
| Hedonic value (5)                | 0.655 | 0.615 | 0.472 | 0.677 | 0.867 |       |       |       |
| Altruistic value (6)             | 0.676 | 0.733 | 0.438 | 0.827 | 0.658 | 0.854 |       |       |
| Social value (7)                 | 0.539 | 0.471 | 0.440 | 0.717 | 0.560 | 0.697 | 0.901 |       |
| Well-being (8)                   | 0.666 | 0.582 | 0.477 | 0.708 | 0.711 | 0.639 | 0.574 | 0.821 |

**Table 8:**  $R^2$  - Values of the Endogenous Latent Variables

| Construct        | $R^2$ |
|------------------|-------|
| Functional value | 0.580 |
| Hedonic value    | 0.455 |
| Altruistic value | 0.538 |
| Social value     | 0.193 |
| Well-being       | 0.603 |

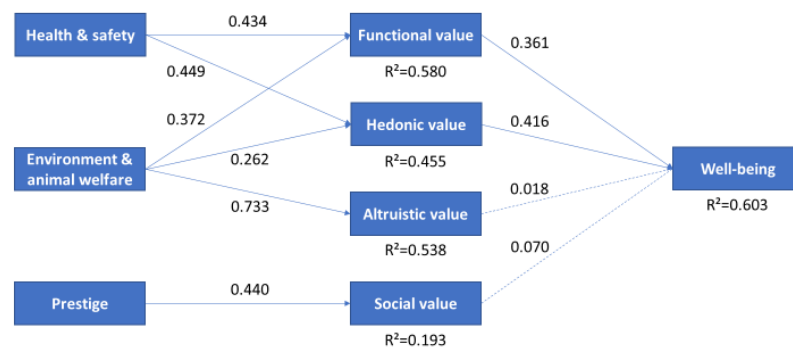
**Table 9:** Significance Testing of the Structural Model Path CoefficientsNote: \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ 

| Path                                            | Path coefficients | t-values  | p-values | Hypothesis         |
|-------------------------------------------------|-------------------|-----------|----------|--------------------|
| Health & safety → Functional value              | 0.434             | 5.154***  | <0.0001  | H2.1a supported    |
| Health & safety → Hedonic value                 | 0.449             | 5.042***  | <0.0001  | H2.1b supported    |
| Environment & animal welfare → Functional value | 0.372             | 4.259***  | <0.0001  | H2.2a supported    |
| Environment & animal welfare → Hedonic value    | 0.262             | 2.957***  | 0.003    | H2.2b supported    |
| Environment & animal welfare → Altruistic value | 0.733             | 18.701*** | <0.0001  | H2.2c supported    |
| Prestige → Social value                         | 0.440             | 7.112***  | <0.0001  | H2.3 supported     |
| Functional value → Well-being                   | 0.361             | 3.927***  | 0.0001   | H3.1 supported     |
| Hedonic value → Well-being                      | 0.416             | 6.127***  | <0.0001  | H3.2 supported     |
| Altruistic value → Well-being                   | 0.018             | 0.199     | 0.843    | H3.3 not supported |
| Social value → Well-being                       | 0.070             | 0.935     | 0.350    | H3.4 not supported |

in terms of environmental friendliness & animal welfare in comparison to the control group. In other words, the label is used as a cue for consumers' quality perception leading to positive inferences about the product's environmental performance. Of all quality dimensions studied, this showed the most significant difference between the two groups, indicating consumers' strong cognitive links between the organic la-

bel and ecological issues. Furthermore, the label leads to higher ratings of the food offered in terms of prestige. Referring to a good's superior image, prestige has already been identified as an important quality dimension for example in a study by Brucks et al. (2000). As organics are generally considered to be premium products that attract higher socio-economic classes (e.g. Harper and Makatouni, 2002; Hill and





**Figure 4:** Structural Results of the Final Model

Lynchehaun, 2002) this inference seems to be consequential. The organic label is the most successful eco-label on the market and is perceived positively by customers (Howard and Allen, 2006). Moreover, it has been postulated before that the organic logo is similar to a brand (MINTEL, 2000) and activates the same brain sections like luxury goods (Linder et al., 2010). The results of this survey show that the label positively biases customers' quality evaluations through inferences about its exclusiveness and positive image. This is an interesting finding that might play an important role in the consumption of organic products as the use of goods is always related to symbolic issues. Prestige has so far not received much of attention in the organic food literature, but the findings indicate that it is actually quite influential in consumers' quality evaluations and should thus be considered in future studies.

In sum, the findings of this survey are in line with previous works on the quality perceptions of organic food and show that the label's halo effect also applies in a canteen setting. In fact, the rather low means in the control group for the quality dimensions under study, ranging from 2.89 for environmental friendliness & animal welfare to 3.86 for prestige demonstrate that consumers in general perceive the normal, average canteen food as being of low quality. It has been argued before, that customers expect less from meals in canteens in terms of quality and taste in comparison to food cooked at home or in a restaurant (Price et al., 2016). In this case, labels can be used as cues indicating the product's hidden characteristics (Fernqvist and Ekelund, 2014) leading to higher quality evaluations of credence attributes like health & safety or animal welfare.

Last but not least, the results of the t-test further reveal significant differences in the evaluations of the other constructs in the model. On average, participants in the organic setting indicate a higher value-in-use in all four dimensions under study and eventually also a superior state of post-prandial well-being. Therefore, in a second step, the relationships between the constructs in the model were analysed as it was hypothesized that the positive quality perceptions evoked by the label are causative for consumers higher

value-in-use and well-being evaluations.

## 5.2. Links between Quality Perceptions and Value-in-Use

Previous research has shown, that when eating food consumers obtain value from its healthiness, a pleasurable experience or the feeling to contribute towards environmental protection (Wier et al., 2008). In means-end chain theory, value is believed to stem from quality attributes (Zeithaml, 1988). Hence, it was decided to take a closer look at the links between different value dimensions and consumers' quality perceptions.

The results of the SEM-modelling show that health & safety aspects of food products are the more important drivers of consumers' functional value-in-use than environmental friendliness & animal welfare. This was to be expected, as this type of value, classified as self-oriented, is primarily sought for personal benefits (Holbrook, 2006). It is also in line with Grunert (1995) who posited that the functional value of food stems from its nutritional value and healthiness. Moreover, the results of this study show that naturalness such as the absence of chemicals or artificial ingredients is important for customers, providing them with functional benefits. Natural products are generally considered healthier as consumers link the use of chemicals and food additives to unknown long-term health effects (Lockie et al., 2004; Magnusson et al., 2003). As organics are considered superior in terms of health & safety due to the halo effect, their consumption provides customers with higher levels of functional value. This is in accordance with previous findings of consumers trying to achieve the higher goal of a healthy life through eating organic products (Vega-Zamora et al., 2014).

The environmental quality of food is likewise positively related to functional value. It has been argued before that the use of less pesticides and fertilizers is not only beneficial for the environment but also for the individual, due to functional health benefits especially in the case of fruits and vegetables (Padel and Foster, 2005; Wandel and Bugge, 1997). The data of this survey support the hypothesized link between functional value and environmental friendliness, thereby contradicting findings of lower functional value for eco-friendly

goods in other product categories like household cleaners as postulated by Green and Peloza (2011). Food products that are considered environmentally friendly are actually perceived advantageous in functional value by consumers due to the health benefits associated with the naturalness and non-use of pesticides and chemicals.

In addition, animal welfare issues are found to be positively related to functional value, driven by the notion of "you are what you eat" (Makatouni, 2002, p. 351). The use of growth hormones or antibiotics in intensive farming to produce meat fast and cheap is opposed by a growing number of costumers (Janet Eastwood, 1995). At the same time, specialised animal welfare labels as currently discussed in Germany find high levels of support among the general population (Howard and Allen, 2006). Customers see a clear link between happy, healthy livestock and the quality of their products (Harper and Makatouni, 2002). Hence, it can be concluded that consumers are increasingly interested in the rearing conditions of animals and that they link those to functional benefits when consuming. Summing up, the functional value of food is found to be driven by a product's quality perception of health & safety as well as environmental friendliness & animal welfare.

Hedonic value is derived from feelings and affections in a consumption experience (Sheth et al., 1991). It seems reasonable that health & safety aspects, like the use or non-use of specific ingredients impacts on how people feel about products (Lee and Yun, 2015; Padel and Foster, 2005). Long lists of e-numbers for instance will probably lead to disapproval among consumers. It has been argued before that naturalness of food products leads to pleasurable feelings and emotional appeal (Lockie et al., 2004; Vega-Zamora et al., 2014). Baker et al. (2004) for example showed that the use of natural ingredients is related to higher pleasure and enjoyment when consuming. This research extends those findings by showing that the health & safety aspects of food products directly and significantly impact on consumers hedonic value-in-use when eating them. As food in the normal canteen is believed to include many artificial ingredients and is rated rather low in terms of healthiness, it seems only consequential that in those canteens customers on average have less fun and experience lower enjoyment when eating there. Hence, the higher hedonic satisfaction in the experimental setting is the consequence of the use of an organic label that leads to positive inferences about the food's naturalness and healthfulness.

Besides, environmental friendliness & animal welfare is also significantly related to hedonic value. Customers seem to enjoy eating food that stems from animal friendly production and is less harmful for the environment. Some authors have postulated before that people engage in ethical activities for their personal contentment (De Young, 1996) and that they feel good using eco-friendly products (Pickett-Baker and Ozaki, 2008). Consumers of organic produce for instance have been shown to experience emotional satisfaction in consumption situations, driven by the belief to support better animal rearing conditions (Harper and Makatouni, 2002).

Hence, it can be summarized that the perception of food products as being environmentally friendly leads to higher hedonic ratings, however only to a lesser extent than health & safety aspects, which seem to be predominant for both functional and hedonic value evaluations.

With regards to altruistic value, environment & animal welfare is highly influential with a path coefficient above 0.7. This is the highest link between any two constructs in the model. In addition, this single factor is able to explain 53,8% of the variance in altruistic value, showing that the link between the two constructs is very strong. This seems reasonable, since this type of value by definition stems from the welfare of others and encompasses ethical issues (Holbrook, 2006; De Young, 1996). It has been argued before that consuming in line with one's moral principles provides customers with a "warm glow" (Andreoni, 1990, p. 464). In case of organic products, Dean et al. (2008) identified the feeling to contribute to something better as an important purchasing motive. This research shows, that when products are considered superior in terms of environmental quality & animal welfare, like in case of organics, they provide consumers with higher levels of altruistic value.

Last but not least, social value was analysed in this study. It was hypothesized that food that is considered prestigious leads to a higher status and self-image when consuming, thus providing social value (Holbrook, 1999). The results of this study show support for this notion. When people perceive products to be special and liked by others, they attribute higher social value in a usage situation. As Levy (1959) so famously noted, products carry symbolic meaning and can be considered representatives of a specific lifestyle. Moreover, it has been argued that food choices and the adoption of a particular diet like veganism is also driven by the desire to be associated with a particular group (Bublitz et al., 2010). Consequently, food can provide social value, especially in the case of organics that are considered to be status symbols enhancing social standing (Barrena and Sanchez, 2010).

### 5.3. Impact on Post-Prandial Well-Being

In means-end chain theory, values can be defined as specific end-states of existence that serve as guiding principles in life (Rokeach, 1973). Following the growing stream of research, well-being was analysed as an important goal consumers try to achieve through their food consumption (e.g. Boelsma et al., 2010; Bublitz et al., 2013). Previous research showed that when prompted to think about food well-being, customers referred to health aspects, nutritional value or naturalness (Ares et al., 2014). Yiridoe et al. (2005) postulated that the quality attributes of food products represent input factors in consumers' demand function for well-being. However, this research followed a different argumentation. In line with means-end chain theory (Gutman, 1982), it was argued that food quality only indirectly influences consumer well-being through the benefits associated with its consumption, which are reflected in value-in-use (Woodruff, 1997).

The results of this research show support for this relation. Functional and hedonic value are found to both sig-

nificantly contribute to consumer well-being. This consistent with [Bublitz et al. \(2013\)](#) who contended that FWB stems among others from functional and hedonic consequences. Experiencing hedonic satisfaction in a consumption situation seems to be beneficial for consumers' welfare, extending towards their evaluation of their post-prandial state of being. Similarly, when food is considered to provide functional health benefits, cognitive inferences lead to higher ratings of subjective well-being after consumption. Both links were highly significant, indicating their influential effect in the model.

Besides, it was hypothesized that social value would also positively impact on well-being since social aspects are generally believed to be a highly influential factor of the construct (e.g. [Block et al., 2011](#); [Grant et al., 2007](#)). However, the path coefficient is found to be non-significant. The reason might be that in this research only status and prestige were considered as input factors of social value. Other aspects like belongingness, comfort, affection or friendship that might be highly influential for social value especially in the work place were not included. This could explain why no relation to well-being could be substantiated by the data.

Equally, the data of this research show that the link between altruistic value and well-being was not significant. While a food's quality evaluations in terms of environmental friendliness lead to higher ratings of altruistic value, this in turn does not impact on post-prandial well-being. Hypothesis 3.4 could hence not be supported. While contribution to society has been found to be an important constituent of the construct ([Marks and Shah, 2004](#)), it seems like the other value dimensions are just more influential for consumers' evaluation of subjective well-being.

Taken together, it can be concluded that the provision of healthy food that satisfies employees' functional needs and gives them hedonic pleasure, ultimately benefits their well-being. [Haugaard et al. \(2016\)](#) and [Price et al. \(2016\)](#) were among the first to propose that workplace lunches might influence employee well-being. The data of this research empirically supported this notion. The findings could therefore be of interest for practitioners and HR-managers intending to enhance employee well-being.

## 6. Practical Implications

Food well-being is an interesting topic for politicians and practitioners around the world. Promoting sustainable food is regarded as being important in the face of global challenges like the growing number of civilization diseases ([Larceneux et al., 2012](#)). As much of people's life is spent at work, the workplace is considered to be an influential factor in promoting healthy diets ([Price et al., 2016](#)). In line with the findings of [Price et al. \(2016\)](#), this research however shows that customers generally evaluate the food in a normal, average canteen as being of low quality in terms of healthiness, environmental friendliness, animal welfare and prestige. At the same time, consumers desire information on production related aspects, for example due to food safety reasons. They

want to eat healthy meals and experience hedonic pleasure from its consumption ([Zanoli and Naspetti, 2002](#)). Hence, the promotion of credence attributes of food products could play a major role in enhancing quality and value perceptions and eventually contribute to more sustainable consumption patterns ([Lee and Hwang, 2016](#)).

Besides offering fitness classes or flexible working hours ([Evans and Prilleltensky, 2007](#)), HR-managers should thus also consider investments into on-site food offerings to enhance the well-being of their workforce. One option might be to offer organic food in the canteen. [Apaolaza et al. \(2018\)](#) were the first to empirically attest a positive link between organic products and consumer well-being. They attributed this to the label's halo effect but asked for future contributions to analyse the mechanism behind it. This research shows that using eco-labels in a canteen leads to positive inferences about the quality of the food offered, a higher value-in-use and ultimately enhanced well-being. Other authors have postulated before that employee well-being is positively related to job performance, motivation and absenteeism (e.g. [Diener, 2006](#); [Tenney et al., 2016](#)). Accordingly, providing high-quality food in a canteen might prove beneficial not only for the employees but likewise for the company.

So far however, availability of organic food in canteens is rather limited (e.g. [Lockie et al., 2002](#); [Zanoli and Naspetti, 2002](#)). The latest [Ökobarometer \(2017\)](#) revealed that despite favourable attitudes among consumers towards organic meals in the workplace, they were only provided by a minority, especially in rural areas. In fact, only 3% of all canteens have been found to use only organic ingredients ([Ökobarometer, 2017](#), p. 18). Besides, higher prices of those options have been identified as purchasing barriers (e.g. [Magnusson et al., 2001](#); [Zanoli and Naspetti, 2002](#)). Offering organic meals in canteens would hence require some investments from companies, probably including subsidies for those meals in order not to inflate prices for the employees. Nevertheless, this investment could be valuable due to positive effects on job performance and lower costs for healthcare and absent days.

In this study, prestige is found to be a less important construct. While it leads to higher ratings of social value, no link to well-being can be substantiated. Therefore, food providers should focus on highlighting quality aspects linked to health & safety and the environment. They could for instance highlight the naturalness of their offer thereby addressing emotional desires and food safety concerns simultaneously ([Lockie et al., 2004](#)). When employees think they can realise functional and hedonic benefits through the consumption of a product, they will not only benefit from enhanced post-prandial well-being but also be more likely to buy it in the first place ([Barrena and Sanchez, 2010](#)). Communicating the benefits of organic food in terms of quality and value may thus prove successful in justifying price premiums and making more people choose those options ([Padel and Foster, 2005](#)).

## 7. Conclusion

Having discussed the practical implications of this research, the paper ends with some concluding remarks. First up, a short summary will highlight the most important results of this project. Thereafter, the limitations of the survey will be discussed leading to recommendations for future research.

### 7.1. Summary of the Findings

This research contributes to the growing stream of literature on food well-being. The aim of this paper was to analyse how consumer perceptions of food quality impact on value-in-use evaluations and well-being. Specifically, it tested the organic label halo effect in an experimental canteen setting. For that purpose, a quantitative study was conducted with two groups, one being treated with the label.

The halo effect was verified using a two tailed t-test. It was shown that participants in the organic setting evaluated the food offer significantly higher in terms of health & safety, environmental quality & animal welfare as well as prestige. As both groups received exactly the same stimuli besides the organic logo, the differences in quality perceptions can be attributed to the label. Hence, it can be concluded that the label acts as a cue that positively biases consumers' quality evaluations of credence attributes. Thus, this research indicated that the organic label halo effect also applies in a canteen setting.

To examine the hypothesized relationships between quality, value and well-being, structural equation modelling (SEM) was used. The results of the SEM revealed that the quality aspects under study were significantly related to the value-in-use dimensions. Functional and hedonic value perceptions were both driven by health & safety as well as environmental friendliness & animal welfare. The latter construct further contributed significantly to altruistic value. Finally, prestige was found to enhance consumers' social value.

The hypothesized link between value-in-use and well-being could only be partially supported. While altruistic and social value did not impact on consumers evaluations of post-prandial well-being, functional and hedonic aspects were found to be highly significant.

Taken together, it can be concluded that higher quality perceptions of food products lead to higher evaluations of value-in-use and subjective post-prandial well-being. As organic products were considered superior in terms of all quality dimensions included in this study, offering those options in a canteen might prove beneficial for companies as it ultimately enhances the well-being of their workforce.

### 7.2. Limitations and Recommendations for Future Research

Like every study, this research has several limitations. First up, as self-selection and convenience sampling were deployed, the answers cannot be considered representative for the overall population. In fact, the topic of the study might

have attracted a specific group of respondents, thereby biasing the results. Hence, while the findings were insightful, the limited representativeness of the sample restricts the possibility to draw reliable conclusions and to generalise the findings.

Moreover, a different study design could have been applied. Instead of doing an online experiment, a real-life study would have been an alternative option to assess consumers' quality and value perceptions in a canteen setting. However, in real-life studies, contextual factors such as ambience or social context may influence consumer perception of a meal experience (Haugaard et al., 2016). Therefore, an online survey was considered to be more adequate to test the impact of the organic label halo effect.

In former studies on food quality, further dimensions such as sensory appeal, convenience or political issues like fairness were included (e.g. Lindeman and Väänänen, 2000; Steptoe et al., 1995). However, given the constraints of this research, it was necessary to focus on a few specific constructs. Health & safety as well as environmental friendliness & animal welfare are among the two most frequently cited quality dimensions of food products (e.g. Magnusson et al., 2003; Makatouni, 2002; Wandel and Bugge, 1997). Hence, they were chosen for this analysis. Besides, prestige was included as an interesting dimension, that, so far, has not received much attention in the scientific food literature.

Well-being was analysed as an important end-goal consumers try to achieve through their consumption choices. However, measuring it in an experimental setting is difficult and may be skewed as it is a long-term concept (Apao-laza et al., 2018). Therefore, to actually prove the impact of food consumption on perceived well-being, a long-term study would be necessary comparing one group of people following an organic diet with a respective control group. This would also allow to further assess the question of whether organic products are actually healthier or whether that link simply stems from a purely psychological effect of the label (Apao-laza et al., 2018).

Given the time constraints of this project, no further examination of the barriers that are commonly associated with organic products could be made. It has been recognized before, that while customers overwhelmingly state favourable attitudes towards organic goods, they do not always make the corresponding purchasing decisions (e.g. Wandel and Bugge, 1997). This discrepancy due to conflicts between environmental benefits and convenience or price has been termed attitude-behaviour gap and is regularly recognized in the organic literature (e.g. Magnusson et al., 2001; Padel and Foster, 2005). Consequently, it is questionable whether in real life situations organic options offered in canteens will be equally perceived as in the experimental setting of this study.

Furthermore, mistrust in eco-labels is an important barrier of sustainable consumption (e.g. Drexler and Fiala, 2018; Padel and Foster, 2005). Cases of mislabelling and non-confirmation to standards reported by the media have contributed to some scepticism among consumers (Yiridoe



et al., 2005). In other words, the organic label halo effect might interact with customers' level of trust in compliance and monitoring of organic standards. Accordingly, future studies should look at how trust moderates the relationship between the label and consumers' quality perceptions.

Other factors that could be of interest for future research are health consciousness and environmental concern. Both are likely to moderate the relationships between the label, consumers' quality evaluations and their value-in-use. People scoring high on environmental concern for instance might rate organic food higher in terms of environmental quality and exhibit higher levels of hedonic and altruistic value when eating those products. Similarly, health conscious individuals might experience higher functional value when consuming organics if, due to the halo effect, they rate them as being high in nutrients or low in calories.

Further demographical data may prove helpful in order to better understand organic food consumption behaviour. Johnston et al. (2011) for example found that individuals with higher education and income were more engaged in ethical consumption. Besides, gender or age have been found to moderate those relations (Ökobarometer, 2017). These factors are likely to impact on the links between quality and value and should thus be considered in future research on that topic.

Last but not least, despite careful conduction of this research, the analysis and presentation of the data may have unintentionally been biased by the researcher, who is particularly interested in the topic of sustainable consumption. This was dealt with by closely examining the findings in comparison to existing literature and acting with due diligence at all stages of the research process.



## References

- Aaker, D. A. Managing Brand Equity, New York: The Free Press, 1991.
- Abbott, L. Quality and Competition, New York: Columbia University Press, 1955.
- Ajzen, I. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211, 1991.
- Andreoni, J. Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, 100(401):464–477, 1990.
- Apaolaza, V., Hartmann, P., Echebarria, C., and Barrutia, J. M. Organic label's halo effect on sensory and hedonic experience of wine: A pilot study. *Journal of sensory studies*, 32(1), 2017.
- Apaolaza, V., Hartmann, P., D'Souza, C., and López, C. M. Eat organic–feel good? the relationship between organic food consumption, health concern and subjective wellbeing. *Food quality and preference*, 63:51–62, 2018.
- Apetito. Apetito menu charts 2017, 2018. URL [https://www.apetito.de/presse/2018/05/apetito\\_menuecharts](https://www.apetito.de/presse/2018/05/apetito_menuecharts). 20.05.2018.
- Ares, G., De Saldamando, L., Giménez, A., and Deliza, R. Food and wellbeing. towards a consumer-based approach. *Appetite*, 74:61–69, 2014.
- Baker, S., Thompson, K. E., Engelken, J., and Huntley, K. Mapping the values driving organic food choice: Germany vs the uk. *European journal of marketing*, 38(8):995–1012, 2004.
- Barrena, R. and Sanchez, M. Frequency of consumption and changing determinants of purchase decision: from attributes to values in the organic food market. *Spanish Journal of Agricultural Research*, 8(2):251–272, 2010.
- BLE Bundesanstalt für Landwirtschaft und Ernährung. Bio-siegel, 2016. URL [https://www.oekolandbau.de/fileadmin/redaktion/Bildarchiv/Bio-Siegel/user\\_upload/Dokumente/Broschueren/flyer\\_biosiegel\\_din\\_a5\\_2016\\_web.pdf](https://www.oekolandbau.de/fileadmin/redaktion/Bildarchiv/Bio-Siegel/user_upload/Dokumente/Broschueren/flyer_biosiegel_din_a5_2016_web.pdf). 27.06.2018.
- Block, L. G., Grier, S. A., Childers, T. L., Davis, B., Ebert, J. E., Kumanyika, S., Laczniak, R. N., Machin, J. E., Motley, C. M., Peracchio, L., et al. From nutrients to nurturance: A conceptual introduction to food well-being. *Journal of Public Policy & Marketing*, 30(1):5–13, 2011.
- Boelsma, E., Brink, E. J., Stafleu, A., and Hendriks, H. F. Measures of post-prandial wellness after single intake of two protein–carbohydrate meals. *Appetite*, 54(3):456–464, 2010.
- Brucks, M., Zeithaml, V. A., and Naylor, G. Price and brand name as indicators of quality dimensions for consumer durables. *Journal of the academy of marketing science*, 28(3):359–374, 2000.
- Bruns, K. and Jacob, F. Value-in-use: Antecedents, dimensions, and consequences. *Marketing ZFP*, 38(3):135–149, 2016.
- Bublitz, M. G., Peracchio, L. A., and Block, L. G. Why did i eat that? perspectives on food decision making and dietary restraint. *Journal of Consumer Psychology*, 20(3):239–258, 2010.
- Bublitz, M. G., Peracchio, L. A., Andreasen, A. R., Kees, J., Kidwell, B., Miller, E. G., Motley, C. M., Peter, P. C., Rajagopal, P., Scott, M. L., et al. Promoting positive change: Advancing the food well-being paradigm. *Journal of Business Research*, 66(8):1211–1218, 2013.
- Caswell, J. A. and Mojduszka, E. M. Using informational labeling to influence the market for quality in food products. *American Journal of Agricultural Economics*, 78(5):1248–1253, 1996.
- Chin, W. W. How to write up and report pls analyses. In *Handbook of partial least squares*, pages 655–690. Springer, 2010.
- Collins, A. M. and Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological review*, 82(6):407–428, 1975.
- Cox, D. F. Risk taking and information handling in consumer behavior. *Cambridge: Harvard University Press*, 1967.
- Darby, M. R. and Karni, E. Free competition and the optimal amount of fraud. *The Journal of law and economics*, 16(1):67–88, 1973.
- De Young, R. Some psychological aspects of reduced consumption behavior: The role of intrinsic satisfaction and competence motivation. *Environment and behavior*, 28(3):358–409, 1996.
- Dean, M., Raats, M. M., and Shepherd, R. Moral concerns and consumer choice of fresh and processed organic foods 1. *Journal of Applied Social Psychology*, 38(8):2088–2107, 2008.
- Dictionary, O. E. Volume XX, Wave-Zyxt, 2nd edition, Oxford: Clarendon Press., 1989.
- Diener, E. Guidelines for national indicators of subjective well-being and ill-being. *Applied research in quality of life*, 1(2):151–157, 2006.
- Drexler, D. and Fiala, J. The effect of organic food labels on consumer attention. *Journal of Food Products Marketing*, 24(4):441–455, 2018.
- EC. Council Regulation (EC) No 834/2007 of 28 June 2007 on organic production and labelling of organic products and repealing Regulation (EEC) No 2092/91, Official Journal of the European Union, L 189, p. 1–23, 2007. URL <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:189:0001:0023:DE:PDF>. 26.06.2018.
- Evans, S. D. and Prilleltensky, I. Youth and democracy: Participation for personal, relational, and collective well-being. *Journal of community psychology*, 35(6):681–692, 2007.
- Fernqvist, F. and Ekelund, L. Credence and the effect on consumer liking of food—a review. *Food Quality and Preference*, 32:340–353, 2014.
- Golder, P. N., Mitra, D., and Moorman, C. What is quality? an integrative framework of processes and states. *Journal of Marketing*, 76(4):1–23, 2012.
- Grankvist, G., Dahlstrand, U., and Biel, A. The impact of environmental labelling on consumer preference: Negative vs. positive labels. *Journal of Consumer Policy*, 27(2):213–230, 2004.
- Grant, A. M., Christianson, M. K., and Price, R. H. Happiness, health, or relationships? managerial practices and employee well-being tradeoffs. *Academy of management perspectives*, 21(3):51–63, 2007.
- Green, T. and Peloza, J. How does corporate social responsibility create value for consumers? *Journal of consumer marketing*, 28(1):48–56, 2011.
- Grönroos, C. Service logic revisited: who creates value? and who co-creates? *European business review*, 20(4):298–314, 2008.
- Grunert, K. G. Food quality: A means-end perspective. *Food Quality and Preference*, 6(3):171–176, 1995.
- Gutman, J. A means-end chain model based on consumer categorization processes. *Journal of marketing*, 46(2):60–72, 1982.
- Hair Jr., J. F., Hult, G. T. M., Ringle, C., and Sarstedt, M. A primer on partial least squares structural equation modeling (PLS-SEM), Los Angeles [u.a.]: Sage Publications, 2014.
- Harper, G. C. and Makatouni, A. Consumer perception of organic food production and farm animal welfare. *British Food Journal*, 104(3/4/5):287–299, 2002.
- Haugaard, P., Stancu, C. M., Brockhoff, P. B., Thorsdottir, I., and Lähteenmäki, L. Determinants of meal satisfaction in a workplace environment. *Appetite*, 105:195–203, 2016.
- Hill, H. and Lynchehaun, F. Organic milk: attitudes and consumption patterns. *British Food Journal*, 104(7):526–542, 2002.
- Hoefkens, C., Verbeke, W., Aertsens, J., Mondelaers, K., and Van Camp, J. The nutritional and toxicological value of organic vegetables: Consumer perception versus scientific evidence. *British Food Journal*, 111(10):1062–1077, 2009.
- Holbrook, M. B. The nature of customer value: an axiology of services in the consumption experience. *Service quality: New directions in theory and practice*, 21:21–71, 1994.
- Holbrook, M. B. Introduction to consumer value, consumer value: a framework for analysis and research. *Introduction to consumer value. Consumer Value: A Framework for Analysis and Research*. Routledge, London, pages 1–28, 1999.
- Holbrook, M. B. Consumption experience, customer value, and subjective personal introspection: An illustrative photographic essay. *Journal of business research*, 59(6):714–725, 2006.
- Howard, P. H. and Allen, P. Beyond organic: consumer interest in new labelling schemes in the central coast of california. *International Journal of Consumer Studies*, 30(5):439–451, 2006.
- Hughner, R. S., McDonagh, P., Prothero, A., Shultz, C. J., and Stanton, J. Who are organic food consumers? a compilation and review of why people purchase organic food. *Journal of Consumer Behaviour: An International Research Review*, 6(2-3):94–110, 2007.
- Janet Eastwood, P. Farm animal welfare, europe and the meat manufacturer. *British Food Journal*, 97(9):4–11, 1995.
- Johnston, J., Szabo, M., and Rodney, A. Good food, good people: Understanding the cultural repertoire of ethical eating. *Journal of consumer culture*, 11(3):293–318, 2011.
- Kleinaltenkamp, M., Storck, F., Gumprecht, P., and Li, J. The impact of psychological ownership on value in use and relational outcomes. *SMR-Journal of Service Management Research*, 2(2):50–70, 2018.
- Larceneux, F., Benoit-Moreau, E., and Renaudin, V. Why might organic labels fail to influence consumer choices? marginal labelling and brand equity

- effects. *Journal of Consumer Policy*, 35(1):85–104, 2012.
- Lee, H.-J. and Hwang, J. The driving role of consumers' perceived credence attributes in organic food purchase decisions: A comparison of two groups of consumers. *Food quality and preference*, 54:141–151, 2016.
- Lee, H.-J. and Yun, Z.-S. Consumers' perceptions of organic food attributes and cognitive and affective attitudes as determinants of their purchase intentions toward organic food. *Food quality and preference*, 39:259–267, 2015.
- Lee, W.-c. J., Shimizu, M., Kniffin, K. M., and Wansink, B. You taste what you see: Do organic labels bias taste perceptions? *Food Quality and Preference*, 29(1):33–39, 2013.
- Levy, S. J. Symbols for Sale, *Harvard Business Review*, 37 (4), p. 117-124, 1959.
- Lindeman, M. and Väänänen, M. Measurement of ethical food choice motives. *Appetite*, 34(1):55–59, 2000.
- Linder, N. S., Uhl, G., Fliessbach, K., Trautner, P., Elger, C. E., and Weber, B. Organic labeling influences food valuation and choice. *NeuroImage*, 53 (1):215–220, 2010.
- Lockie, S., Lyons, K., Lawrence, G., and Mummery, K. Eating 'green': motivations behind organic food consumption in australia. *Sociologia ruralis*, 42(1):23–40, 2002.
- Lockie, S., Lyons, K., Lawrence, G., and Grice, J. Choosing organics: a path analysis of factors underlying the selection of organic food among australian consumers. *Appetite*, 43(2):135–146, 2004.
- Macdonald, E. K., Kleinaltenkamp, M., and Wilson, H. N. How business customers judge solutions: Solution quality and value in use. *Journal of Marketing*, 80(3):96–120, 2016.
- Magnusson, M. K., Arvola, A., Koivisto Hursti, U.-K., Åberg, L., and Sjöden, P.-O. Attitudes towards organic foods among swedish consumers. *British food journal*, 103(3):209–227, 2001.
- Magnusson, M. K., Arvola, A., Hursti, U.-K. K., Åberg, L., and Sjöden, P.-O. Choice of organic foods is related to perceived consequences for human health and to environmentally friendly behaviour. *Appetite*, 40(2):109–117, 2003.
- Makatouni, A. What motivates consumers to buy organic food in the uk? results from a qualitative study. *British Food Journal*, 104(3/4/5):345–352, 2002.
- Marks, N. and Shah, H. A well-being manifesto for a flourishing society. *Journal of Public Mental Health*, 3(4):9–15, 2004.
- Nations, U. Transforming our world: the 2030 Agenda for Sustainable Development, United Nations, New York, 2015. URL [www.un.org/ga/search/view\\_doc.asp?symbol=A/RES/70/1&Lang=E](http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E). 12.06.2018.
- Nelson, P. Information and consumer behavior. *Journal of political economy*, 78(2):311–329, 1970.
- Olson, J. C. and Jacoby, J. Cue utilization in the quality perception process. *ACR Special Volumes*, 1972.
- Ophuis, P. A. O. and Van Trijp, H. C. Perceived quality: a market driven and consumer oriented approach. *Food quality and Preference*, 6(3):177–183, 1995.
- Padel, S. and Foster, C. Exploring the gap between attitudes and behaviour: Understanding why consumers buy or do not buy organic food. *British food journal*, 107(8):606–625, 2005.
- Pickett-Baker, J. and Ozaki, R. Pro-environmental products: marketing influence on consumer purchase decision. *Journal of consumer marketing*, 25(5):281–293, 2008.
- Porter, M. E. *Competitive strategy: Techniques for analyzing industries and competitors*. Simon and Schuster, 1980.
- Price, S., Hartwell, H., Hemingway, A., and Chapleo, C. Workplace foodservice; perception of quality and trust. *Appetite*, 97:169–175, 2016.
- Renner, B., Sproesser, G., Strohbach, S., and Schupp, H. T. Why we eat what we eat. the eating motivation survey (tems). *Appetite*, 59(1):117–128, 2012.
- Rokeach, M. J. *The Nature of Human Values*, New York: The Free Press., 1973.
- Rust, R. T., Oliver, R. L., et al. Service quality: insights and managerial implications from the frontier. *Service quality: New directions in theory and practice*, pages 1–19, 1994.
- Saunders, M., Lewis, P., and Thornhill, A. *Research methods for business students*, harlow: Pearson education., 2009.
- Schuldt, J. P. and Schwarz, N. The "organic" path to obesity? organic claims influence calorie judgments and exercise recommendations. *Judgment and Decision making*, 5(3):144, 2010.
- Sheth, J. N., Newman, B. I., and Gross, B. L. Why we buy what we buy: A theory of consumption values. *Journal of business research*, 22(2):159–170, 1991.
- Steenkamp, J.-B. E. Conceptual model of the quality perception process. *Journal of Business research*, 21(4):309–333, 1990.
- Stephens, A., Pollard, T. M., and Wardle, J. Development of a measure of the motives underlying the selection of food: the food choice questionnaire. *Appetite*, 25(3):267–284, 1995.
- Sweeney, J. C. and Soutar, G. N. Consumer perceived value: The development of a multiple item scale. *Journal of retailing*, 77(2):203–220, 2001.
- Tenney, E. R., Poole, J. M., and Diener, E. Does positivity enhance work performance?: Why, when, and what we don't know. *Research in Organizational Behavior*, 36:27–46, 2016.
- Thorndike, E. L. A constant error in psychological ratings. *Journal of applied psychology*, 4(1):25–29, 1920.
- Torjusen, H., Lieblein, G., Wandel, M., and Francis, C. A. Food system orientation and quality perception among consumers and producers of organic food in hedmark county, norway. *Food quality and preference*, 12(3):207–216, 2001.
- Vargo, S. L. and Lusch, R. F. Service-dominant logic: continuing the evolution. *Journal of the Academy of Marketing Science*, 36(1):1–10, 2008.
- Vega-Zamora, M., Torres-Ruiz, F. J., Murgado-Armenteros, E. M., and Parras-Rosa, M. Organic as a heuristic cue: What spanish consumers mean by organic foods. *Psychology & Marketing*, 31(5):349–359, 2014.
- Wandel, M. and Bugge, A. Environmental concern in consumer evaluation of food quality. *Food quality and preference*, 8(1):19–26, 1997.
- Wier, M., Jensen, K. O., Andersen, L. M., and Millock, K. The character of demand in mature organic food markets: Great Britain and Denmark compared. *Food Policy*, 33(5):406–421, 2008.
- Woodruff, R. B. Customer value: the next source for competitive advantage. *Journal of the academy of marketing science*, 25(2):139, 1997.
- Yiridoe, E. K., Bonti-Ankomah, S., and Martin, R. C. Comparison of consumer perceptions and preference toward organic versus conventionally produced foods: a review and update of the literature. *Renewable agriculture and food systems*, 20(4):193–205, 2005.
- Zanoli, R. and Naspetti, S. Consumer motivations in the purchase of organic food: a means-end approach. *British food journal*, 104(8):643–653, 2002.
- Zeithaml, V. A. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *Journal of marketing*, 52(3):2–22, 1988.
- Ökobarometer. Bundesministerium für Ernährung und Landwirtschaft, 2017. URL [http://www.bmel.de/SharedDocs/Downloads/Ernaehrung/Oekobarometer2017.pdf?\\_\\_blob=publicationFile](http://www.bmel.de/SharedDocs/Downloads/Ernaehrung/Oekobarometer2017.pdf?__blob=publicationFile). 18.06.2018.



## Charakteristika vs. Carry - Outperformance in Devisenmärkten

Tom O. K. Zeissler

Wirtschaftsuniversität Wien

### Abstract

Die optimale Währungsallokation ist eine der Kernfragen des international agierenden Investors. Ich teste in diesem Zusammenhang empirisch den Mehrwert des Portfoliooptimierungsverfahrens von [Brandt et al. \(2009\)](#) im Devisenkontext und vergleiche das Ergebnis mit einer diversifizierten Carry-Strategie und weiteren Benchmarks. Da im Zuge dieses Verfahrens sogenannte Charakteristika als Inputsignale für die Allokation benötigt werden, selektiere ich auf Basis bereits vorhandener Literatur zur Prädiktabilität von Währungsrenditen die Faktoren Carry, Momentum, realer Wechselkurs, Leistungsbilanzsaldo, Produktionslücke und Volatilität und untersuche die Performance der auf Basis dieser Signale generierten Portfolios über den Zeitraum von Anfang 1990 bis Ende 2017. Ich finde die stärkste Evidenz für Carry, aber auch Momentum, der reale Wechselkurs und der Leistungsbilanzsaldo wirken mitunter als Signal für die Allokation der optimalen Portfoliogewichte interessant. Ebenfalls kann ich feststellen, dass die Profitabilität beinahe aller untersuchten Strategien maßgeblich vom gewählten Anlageuniversum abhängt, wobei bei gegebener Datenverfügbarkeit eine breite Definition mit Industrie- und Entwicklungsländern zu bevorzugen ist. Des Weiteren kann ich teils interessante Diversifikationseigenschaften und positive Wechselwirkungen zwischen den Charakteristika dokumentieren, welche sich durch Kombination mittels des gewählten Optimierungsansatzes realisieren lassen. Auch stelle ich fest, dass Transaktionskosten im Schnitt einen wesentlichen negativen Einfluss auf die Renditen der Strategien haben. Die Sensibilisierung des Verfahrens durch Inklusion eines entsprechenden Transaktionskostenterms führt jedoch zu einer Performance-Regeneration.

**Keywords:** currency investing; portfolio allocation; parametric portfolio policies; carry trade.

### 1. Einleitung

#### 1.1. Eine kurze Geschichte des Carry-Trades und der Effizienz in Devisenmärkten

Devisenmärkte sind für eine Vielzahl von Menschen täglich aus den unterschiedlichsten Anlässen relevant. Sei es für die Planung der nächsten Weltreise oder bei der Entscheidung über eine kapitalintensive Anlage in Assets, welche in Fremdwährungen notiert sind. Währungen dienen dabei stets als eine Art Schmiermittel für die internationale wirtschaftliche Interaktion. Getrieben durch die wachsende Globalisierung seit der zweiten Hälfte des 20. Jahrhunderts gilt der Devisenmarkt mittlerweile als der größte und liquideste Markt der Welt ([Bekaert und Panayotov, 2018](#)) – mit einem durchschnittlichen Umschlag von 5,1 Billionen USD pro Tag ([Bank für Internationalen Zahlungsausgleich, 2016](#)) gemessen über die Periode des Aprils 2016.

Die Themen internationaler Währungsmärkte, wie z.B. die Frage nach deren Effizienz oder die Suche nach potentiellen Prädiktoren von Währungsrenditen, beschäftigen dem-

nach auch die wissenschaftliche Forschung vergleichsweise lange. So sprach sich [Nurkse \(1944\)](#), welcher später am Bretton-Woods-Abkommen beteiligt war, bereits in der ersten Hälfte des 20. Jahrhunderts für ein System mit fixen Kursen aus. Als Argument führte er u.a. an, dass Spekulation destabilisierende Auswirkungen auf die Markteffizienz hat. Dies führe wiederum zu Kosten für Produzenten, Konsumenten und schließlich zu weniger effizienten Allokationsentscheidungen beider Marktteilnehmer. [Friedman \(1953\)](#) hingegen ergriff, bereits vor dem Zusammenbruch des Bretton-Woods-Systems im Jahr 1973, Partei für das heutige flexible Wechselkurssystem und argumentierte, dass sich durch Spekulation von erfahrenen Marktteilnehmern der fundamentale Wert einer Währung einstellen würde. Rückblickend weiß man zwar heute, welches der beiden Wechselkurssysteme sich durchsetzen konnte; die Frage nach der Effizienz des Devisenmarktes sowie die Erklärungen von Anomalien beschäftigen die wissenschaftliche Forschung allerdings weiterhin.

Als klassischer Test auf Markteffizienz wird dabei der Bestand der ungedeckten Zinsparität überprüft. Diese auf

Fisher (1896) zurückgehende ökonomische Gesetzmäßigkeit impliziert basierend auf der Annahme rationaler, risikoneutraler Investoren und einer No-Arbitrage-Überlegung, dass der Zinsunterschied zwischen Ländern eine erwartungstreue Marktschätzung der zukünftigen Wechselkursänderung darstellen sollte. Ob dies empirisch gegeben ist wird meist mittels einer Regression getestet, welche den Zusammenhang zwischen der zukünftigen Währungskursbewegung und dem heutigen Zinsunterschied untersucht.<sup>1</sup> Wenn die ungedeckte Zinsparität Bestand hat, sollte ein Beta-Koeffizient von 1 geschätzt werden, d.h. im Schnitt sollte ein positiver Zinsunterschied eine Abwertung der Hochzinswährung in exakt der Höhe zur Folge haben, sodass die erwarteten Renditen von Einlagen in beiden Währungen identisch sind.

Hansen und Hodrick (1980) sowie Fama (1984) untersuchten dies und kamen zum Ergebnis, dass der Beta-Koeffizient jeder geprüften Währung signifikant von 1 abweicht und darüber hinaus negativ ist. Diese ersten Ergebnisse wurden im Laufe der Zeit durch weitere Studien bestätigt. So konstatierten Froot und Thaler (1990) in einer Meta-Analyse, dass der durchschnittliche Beta-Koeffizient von 75 veröffentlichten Schätzungen -0,88 betrug und keine einzige Studie zu einem Ergebnis gleich oder größer der Nullhypothese von 1 kam. Die aus den Resultaten ableitbare Strategie einer Kreditaufnahme in Niedrigzinswährungen und Anlage in Hochzinswährungen, der sogenannte Carry-Trade, ist demnach eine bekannte Anlagetechnik in der Praxis. Ihre Outperformance ist bereits über einen langen Zeitraum dokumentiert, wie etwa Burnside et al. (2008) mit einem Datensatz von 1976 bis 2007 zeigen.

Zwar herrscht Einigkeit über eine gegebene Abweichung von der ungedeckten Zinsparität in den Daten, bei deren Interpretation gibt es aber verschiedene Ansätze.<sup>2</sup>

Eine Erklärungsmöglichkeit ist, dass Anleger systematische Prognosefehler begehen. Dies kann zwar auf eine Verletzung der Annahme rationaler Investorenerwartungen und Marktineffizienzen hinweisen, muss beides aber nicht zwingend implizieren. So ist bspw. das Vorliegen eines sogenannten Peso-Problems ein möglicher Fall, bei dem die Klassifikation der Regressionsergebnisse als Marktineffizienz voreilig erscheint. Dieser oftmals Nobelpreisträger Milton Friedman zugesprochene Ausdruck bezeichnet die Problematik, dass der Markt bei seiner Zukunftsprognose die Wahrscheinlichkeiten wichtiger Ereignisse einpreisen muss, welche bis dato selten oder überhaupt nicht in einem Datensatz aufgetreten sind (Obstfeld, 1989). Mussa (1982) argumentiert in diesem Zusammenhang, dass Inflationsraten eine schiefe Wahrscheinlichkeitsverteilung besitzen, da sich die Inflation in der Regel eine geraume Zeit innerhalb eines Zielkorridors bewegt, bis es vereinzelt zu einer Hyperinflation kommt. Hierdurch

wird, den Gedanken des Autors folgend, die Inflationserwartungen der Investoren in Zeiten ohne Hyperinflation regelmäßig über der tatsächlich realisierten Inflation liegen. Lewis (1989) stellt in diesem Kontext die mögliche Erklärung vor, dass sich Anleger in einem Lernprozess bzgl. Änderungen der Marktgegebenheiten, wie z.B. die Umstellung vom fixen auf das flexible Währungskurssystem, befinden und daher die Profitmöglichkeiten ex-ante womöglich nicht zu erkennen sind. Sie findet in ihrer empirischen Arbeit für den Zeitraum 1980-85 zwar Belege für diese These, stellt aber bereits zum damaligen Zeitpunkt fest, dass die Abweichungen zu persistent sind, um rein durch diese Erklärung gerechtfertigt zu werden. In einer aktuelleren empirischen Arbeit betrachtet Jurek und Jakub W. (2010) eine Strategie, welche das „Peso-Risiko“ des Carry-Trades mittels Optionen absichert. Er kommt zu dem Schluss, dass dieses zwar eine Rolle spiele, allerdings nicht vollständig die Renditen des Carry-Trades erklärt.

Einen weiteren Erklärungsansatz der gegebenen Abweichung von der ungedeckten Zinsparität in den Daten stellt eine Knappheit in profitsuchendem spekulativen Kapital dar, wie bspw. Jylhä und Suominen (2011), Barroso und Santa-Clara (2015) sowie Filipe und Suominen (2014) argumentieren. Diese Knappheit führt dazu, dass Anomalien nicht zwangsweise vollständig durch Arbitrage verschwinden. Jylhä und Suominen (2011) zeigen dabei empirisch, dass Hedge Fonds-Kapitalströme, welche stellvertretend für die Menge an spekulativem Kapital im Markt stehen, signifikante Auswirkungen auf den Währungskurs haben. Als weiterer Punkt dieser These wird u.a. angeführt, dass Zentralbanken als wesentliche Teilnehmer des Devisenmarktes nicht profitorientiert agieren (Taylor, 1982).

Als dritter Erklärungsansatz gilt wie in anderen Asset-Klassen die Theorie zur Existenz zeitvariierender Risikoprämien. Diese besagt, dass unter der Annahme von risikoaversen Investoren und nicht vollständig diversifizierbarem Risiko im Devisenmarkt der Zinsunterschied zwischen Ländern nicht mehr als eine reine erwartungstreue Marktschätzung der zukünftigen Wechselkursänderung interpretiert werden kann. Vielmehr muss er als Summe einer Marktschätzung der zukünftigen Wechselkursänderung sowie einer Risikoprämie gesehen werden (siehe bspw. Fama (1984) und Korajczyk (1985)). Unter Beibehaltung der Annahme rationaler Markterwartungen würde demnach ein Beta-Koeffizient ungleich 1 bedeuten, dass Zinsbewegungen in Zusammenhang zu Änderungen in dieser zeitvariierenden Risikoprämie stehen. Wie Fama (1984) bemerkt, hat ein Koeffizient kleiner 0, wie er im Großteil der Studien konstatiert wurde, in diesem Kontext die Implikation, dass die Varianz der Risikoprämie tatsächlich größer als die Varianz von sowohl der erwarteten Kursabwertung als auch dem Zinsunterschied sein muss. Fama (1984) schlussfolgert darüber hinaus auf eine negative Korrelation zwischen der erwarteten Kursabwertung und der Risikoprämie.

Die sich aus diesem Paradigma ableitende und herausfordernde Aufgabe besteht darin, die besagte nicht beobachtbare Risikoprämie konsistent und robust mit einem überprüfba-

<sup>1</sup>In einer alternativen Version des Tests untersuchen Forscher bisweilen auch den Zusammenhang zwischen der zukünftigen Währungskursbewegung und der heutigen Forward-Basis, siehe hierzu bspw. Froot und Thaler (1990).

<sup>2</sup>Zur Gewinnung eines Überblicks kann bspw. auf Froot und Thaler (1990) oder Burnside et al. (2011) verwiesen werden.



ren Risikofaktor zu verknüpfen. Als ökonomische Erklärungen für solche Risikoprämien gibt es in der Literatur zahlreiche Ansätze. Lustig et al. (2011) zeigen, dass das Risiko von Carry-Trades über alle Währungen hinweg nicht vollständig diversifizierbar ist. Gemäß dieser Überlegung bilden sie einen systematischen Risikofaktor, welcher hilft die Carry-Prämie zu erklären. Brunnermeier et al. (2008) hingegen sehen den möglichen Grund in Liquiditätsengpässen, welche Spekulanten im Zuge höherer Volatilität und eintretender Margin-Calls zum Glattstellen ihrer Carry-Trades zwingt. Sie weisen damit auf ein wesentliches Crash-Risiko des Carry-Trades hin.

Diverse weitere Ansätze und vorgeschlagene Risikofaktoren mit Einfluss auf die zukünftigen Renditen von Währungsinvestments wurden bereits in der Literatur diskutiert. Einige von ihnen werden im weiteren Verlauf als Charakteristika bei der Portfoliooptimierung in den empirischen Teil dieser Arbeit einfließen. Endgültige Aussagen darüber, ob der Devisenmarkt nun effizient oder ineffizient ist, die Prädiktabilität von Währungsrenditen mit systematischen Prognosefehlern von Investoren bzw. der Knappheit in profitsuchendem spekulativen Kapital zu erklären ist oder zeitvariierende Risikoprämien das adäquate Paradigma sind, stehen immer noch zur Debatte, sind aber nicht Gegenstand der vorliegenden Arbeit.

## 1.2. Ziel, Methoden und Struktur der Arbeit

Das Ziel dieser Arbeit ist zu untersuchen, wie die aus dem Portfoliooptimierungsverfahren von Brandt et al. (2009) resultierenden optimalen Portfolios im Devisenkontext im Vergleich zu einer diversifizierten Carry-Strategie und weiteren Benchmarks abschneiden. Dabei folgt die Arbeit im Weiteren der folgenden Struktur: Das zweite Kapitel widmet sich dem methodischen Vorgehen. Zunächst wird der gewählte Optimierungsansatz von Brandt et al. (2009) zu anderen Optimierungsverfahren abgegrenzt. Es folgt die Vorstellung der gewählten Methodologie für die Optimierung in Anlehnung an Barroso und Santa-Clara (2015), welche den selektierten Ansatz bereits im Devisenmarkt angewandt haben. Im letzten Unterkapitel werden die Performance-Benchmarks definiert.

Kapitel 3 beschäftigt sich mit der Selektion der für den gewählten Ansatz nötigen Charakteristika. Im ersten Unterkapitel wird eine Übersicht der betrachteten Charakteristika und Erklärungen zum Selektionsverfahren gegeben, bevor im nächsten Unterkapitel ein Literaturüberblick zu den Charakteristika erfolgt. Ziel hiervon ist es, einen groben Überblick über empirische Ergebnisse bzgl. der prädiktiven Aussagekraft sowie mögliche theoretische Fundierungen dieser Charakteristika zu vermitteln. Im letzten Unterkapitel lege ich im Detail die Methodologie der verwendeten Charakteristika für die weitere Analyse fest.

Anschließend wird im vierten Kapitel Aufschluss über die verwendeten Daten gegeben. Zuerst definiere ich den gewünschten Datensatz und stelle die Datenquellen vor, bevor das weitere Vorgehen bei spezifischen Fragestellungen der Datenverwendung (bspw. Umgang mit fehlenden Daten) geklärt wird. Im letzten Unterkapitel wird der Selektionspro-

zess des finalen Datensatzes skizziert und dieser kurz vorgestellt.

Im fünften Kapitel folgt die Präsentation der Ergebnisse. Der erste Abschnitt erläutert dabei zunächst kurz die Einteilung des Datensatzes sowie die untersuchten optimalen Portfolios. Als nächstes erfolgt die Präsentation der In-Sample-Ergebnisse, bevor die Resultate des Out-of-Sample-Tests dargestellt werden und danach das Kapitel mit einer Kritik am bisherigen Ansatz schließt, welche weitere Analysen in den folgenden Kapiteln motiviert.

Kapitel 6 untersucht in einer Szenarioanalyse, wie sich die Resultate des vorigen Kapitels ändern, wenn ein anderer Grad der relativen Risikoaversion angenommen wird. Ich wähle dabei eine Halbierung und einen extremen Anstieg des in der Standardanalyse angenommenen Aversionsparameters. In Kapitel 7 erweitere ich im Vergleich zum bisherigen Vorgehen die Definition des Währungsuniversums und betrachte, wie sich die Ergebnisse des fünften Kapitels durch die Inklusion der zusätzlichen Daten verändern. Das achte Kapitel berücksichtigt im Kontext der Analyse des vorherigen Kapitels zusätzlich Transaktionskosten.

Kapitel 9 fasst das beschriebene Thema zusammen, konkludiert und schließt die Arbeit ab.

## 2. Methodisches Vorgehen

### 2.1. Abgrenzung des gewählten zu anderen Optimierungsverfahren

Nicht nur die Frage nach der Effizienz von Devisenmärkten, sondern auch die nach der optimalen Zusammenstellung eines Portfolios von Vermögenswerten beschäftigt die wissenschaftliche Literatur schon seit etwa Mitte des 20. Jahrhunderts. Seit Markowitz (1952) mit seinem Ansatz das  $\mu - \sigma$ -Paradigma begründete und auf die Vorteile der Diversifikation aufmerksam machte, wurden zwar diverse andere Portfoliooptimierungsverfahren<sup>3</sup> präsentiert, aber bis heute definiert Markowitz (1952) und insbesondere die Abwägung zwischen erwarteter Rendite und Risiko eine Art „Industriestandard“<sup>4</sup> in der Praxis – obwohl aus dem Ansatz folgende Nachteile durchaus bekannt sind.

So berücksichtigt bspw. das von Markowitz (1952) vorgeschlagene  $\mu - \sigma$ -Paradigma, wie der Name bereits suggeriert, lediglich Erwartungswert und Standardabweichung der Renditeverteilungen und ignoriert höhere Momente sowie etwaige Präferenzen des Investors diesen gegenüber. Wie Brandt (2010) dokumentiert, haben sich diesem Thema bereits diverse wissenschaftliche Arbeiten angenommen und den Ansatz von Markowitz (1952) erweitert.

Ebenfalls kritisch thematisiert wird bisweilen die ökonomische Performance des Ansatzes von Markowitz (1952), bei welchem im ersten Schritt erwartete Renditen und Kovarianz-Matrizen mittels historischer Daten geschätzt und im zweiten Schritt aus diesen Rückschluss auf die optimalen

<sup>3</sup>Brandt (2010) liefert hierzu einen Literaturüberblick.

<sup>4</sup>Zumindest bezeichnet Brandt (2010) das Verfahren als solchen.



Portfoliogewichte gezogen werden. Duchin und Levy (2009) lassen in einem Wettstreit den Ansatz von Markowitz (1952) gegen den Ansatz der sogenannten naiven Diversifikation oder auch  $\frac{1}{n}$ -Ansatz, welcher mitunter dem babylonischen Talmud zugeschrieben wird, im Aktienkontext gegeneinander antreten. Bei diesem Verfahren wird das Vermögen zu gleichen Teilen in jedes Anlagegut des verfügbaren Universums investiert, sodass zumindest initial eine Gleichgewichtung vorliegt. Die Autoren konstatieren, dass der  $\frac{1}{n}$ -Ansatz für Investoren mit kleinen Portfolios und einer geringen Anzahl an gehaltenen Assets vorteilhafter ist, während für große Portfolios der Ansatz von Markowitz (1952) dominiert. Auch Jacobs et al. (2014) vergleichen im Aktiensegment die Performance von diversen Optimierungsverfahren aus der Literatur wie z.B. Markowitz (1952) mit jener von einfachen Heuristiken, u.a. auch dem  $\frac{1}{n}$ -Ansatz. Die Autoren kommen zum Schluss, dass bezogen auf ein globales Aktienuniversum die getesteten Optimierungsverfahren zu keiner Outperformance im Vergleich zu den vergleichsweise einfachen Heuristiken führen. Zu einem ähnlichen Ergebnis kommen auch DeMiguel et al. (2007).

Bekannt ist auch, dass das Verfahren von Markowitz (1952) teils zu unvernünftigen Ergebnissen führt. So beschreiben bspw. Black und Litterman (1992), dass vermehrt hohe Short-Positionen oder, wenn Short-Positionen im Zuge der Optimierung ausgeschlossen werden, „Randlösungen“ das Resultat sind. Mit diesem Ausdruck bezeichnen die Autoren die Tatsache, dass viele Assets des verfügbaren Investmentuniversums nicht im optimalen Portfolio enthalten sind und letzteres eine Neigung zu Vermögenswerten mit geringer Marktkapitalisierung hat. Den Grund hierfür sehen die Autoren sowohl in der Tatsache, dass erwartete Renditen schwer zu schätzen und historische Daten dafür oft eher schlecht geeignet sind, als auch in der hohen Sensitivität des Ansatzes zu diesen Renditen. DeMiguel et al. (2007) machen in diesem Kontext auf eine beachtlich hohe Anzahl an Datenpunkten aufmerksam, die zum Schätzen eines Mittelwert-Varianz-Modells nötig sind, das eine tatsächlich überlegene Out-of-Sample-Performance generiert.

Black und Litterman (1992) präsentieren als Lösungsansatz ein Verfahren, welches die Erwartungshaltung, die sich aus der Schätzung mittels historischer Daten und der Unterstellung eines Marktgleichgewichts<sup>5</sup> ergibt, mit der individuellen Erwartungshaltung des Investors sowie seiner Sicherheit hinsichtlich des Eintritts dieser persönlichen Prognose kombiniert. Hierdurch ergibt sich ein Portfolio, welches als neutralen Punkt, d.h. ohne spezifisch formulierte Erwartungen des Investors, eine Aufstellung nach Marktkapitalisierung aufweist und somit bereits nicht zu besagten unvernünftigen Portfoliogewichten führen sollte. Anschließend wird die Einschätzung auf Basis historischer Marktdaten mit den Zukunftserwartungen des Investors aktualisiert. Somit kommt eine weitere Informationsquelle ins Spiel, die z.B. den rein

historischen Blick durch zukunftsgerichtete Analysen ergänzen bzw. relativieren kann.

Der in der vorliegenden Arbeit verwendete Ansatz von Brandt et al. (2009) versucht ebenfalls die Problematik, welche mit der Schätzung erwarteter Renditen mittels historischer Daten und daraus folgenden Schätzfehlern in Zusammenhang steht, anzugehen. Dafür schlagen sie ein Optimierungsverfahren vor, welches ohne den Schritt der Modellierung von Renditen auskommt. Die optimalen Portfoliogewichte werden stattdessen über die sogenannten „Charakteristika“ und in Relation zu einem Benchmark-Portfolio bestimmt. Diesem Vorgang liegt die Annahme zugrunde, dass die gewählten Charakteristika implizit alle nötigen Informationen zur Evaluierung der bedingten Wahrscheinlichkeitsverteilung der Renditen beinhalten. Demzufolge sind nicht historische Renditen, sondern lediglich die Charakteristika für das Gewicht eines Assets im optimalen Portfolio relevant. Dieser Ansatz führt im Weiteren zu einer starken Reduktion der Dimensionalität des Optimierungsproblems mit einer geringeren Anzahl von zu schätzenden Parametern.

Da die entscheidenden Inputfaktoren des Ansatzes die Charakteristika sind, kann dieser nur Anwendung finden, wenn hiervon brauchbare zur Beschreibung der bedingten Wahrscheinlichkeitsverteilung von Devisenrenditen zur Verfügung stehen. Ich definiere eine brauchbare Charakteristik in diesem Kontext auf Basis ihrer vorhandenen empirischen Dokumentation als relevanter Faktor sowie theoretischen Fundierung in wissenschaftlichen Arbeiten zur Prädiktabilität von Devisenrenditen. Der eingangs erwähnte Carry-Trade ist dieser Definition nach nur eine mögliche Charakteristik, wie in Folge Kapitel 3 zeigt.

## 2.2. Methodologie der Portfoliooptimierung

Die Wahl der Methodologie erfolgt in Anlehnung an Brandt et al. (2009) und Barroso und Santa-Clara (2015).

Dabei liegt zu jedem Zeitpunkt  $t$  die Möglichkeit vor, über FX-Forwards in ein Universum von  $N_t$  Währungen zu investieren, wobei  $N$  für die Gesamtanzahl an unterschiedlichen Währungen im Universum steht. Aus Gründen der Datenverfügbarkeit wird die Perspektive eines US-Investors unterstellt, d.h. der USD wird als die einheimische Währung betrachtet und der Kurs anderer Währungen im Verhältnis zu einem USD ausgedrückt. Demnach bedeutet ein Anstieg des Kurses einen stärkeren USD und eine schwächere Fremdwährung.

Ein Long-Investment in Währung  $i$  zum Zeitpunkt  $t$  führt demnach zu folgender Rendite ausgedrückt in USD in der nächsten Periode:

$$r_{i,t+1} = \frac{F_{i,t:t+1}}{S_{i,t+1}} - 1, \quad (1)$$

wobei  $F_{i,t:t+1}$  den zum Zeitpunkt  $t$  vereinbarten Forward-Kurs von Währung  $i$  mit Fälligkeitszeitpunkt in  $t+1$  und  $S_{i,t+1}$  den Spot-Kurs zum Zeitpunkt  $t+1$  bezeichnet. Die Optimierung erfolgt in der weiteren Analyse wie in Barroso und

<sup>5</sup>Hierbei orientieren sie sich an der Idee des CAPM, wie Sharpe (1964) bzw. Lintner (1965) es beschreiben.

Santa-Clara (2015) mit einer monatlichen Frequenz, d.h. es werden monatliche Daten zu Spot- und 1-Monat-Forward-Kursen herangezogen. Dies impliziert, dass es sich bei Formel (1) um einen um eine monatliche Rendite handelt und zum anderen per se keine Positionen von einem zum folgenden Monat bestehen bleiben. Wie in Barroso und Santa-Clara (2015) wird von etwaigen Sicherheitsforderungen des Brokers abstrahiert. Der Abschluss eines Forward-Kontraktes kann daher als Strategie ohne initiale Kapitalbindung betrachtet werden. Will ein Investor nun in ein Währungsportfolio investieren, sollte er die Portfoliogewichte  $w_{i,t}$  der einzelnen Währungen zum Zeitpunkt der Optimierung so wählen, dass er seinen bedingten erwarteten Nutzen maximiert, welcher wiederum von der Portfoliorendite  $r_{p,t+1}$  abhängt:

$$\max_{\{w_{i,t}\}_{i=1}^{N_t}} E_t[u(r_{p,t+1})] = E_t[u(rf_{US,t:t+1} + \sum_{i=1}^{N_t} w_{i,t} r_{i,t+1})] \quad (2)$$

Wie aus Formel (2) ersichtlich, besteht das betrachtete Gesamtportfolio aus einem 100%igen Investment in die risikolose Geldanlage im Heimatland, hier also der USA, mit der Verzinsung  $rf_{US,t:t+1}$  sowie einem Long-Short-Portfolio aus Forward-Kontrakten. Der Leverage-Effekt  $L_t$  dieses Long-Short-Portfolios wird wie in Barroso und Santa-Clara (2015) als der absolute Betrag definiert, welcher in das riskante Währungsportfolio angelegt ist, gemessen pro investiertem Dollar in der risikolosen Geldanlage:

$$L_t = \sum_{i=1}^{N_t} |w_{i,t}|. \quad (3)$$

Den Umschlag (Turnover)  $T_t$  des besagten Portfolios berechne ich wie in Brandt et al. (2009):

$$T_t = \sum_{i=1}^{N_t} |w_{i,t} - w_{i,t-1}|. \quad (4)$$

Analog zu Barroso und Santa-Clara (2015) werden keine einschränkenden Annahmen in Bezug auf Portfoliogewichte, Leverage oder Turnover getroffen, ich werde aber beide Kennzahlen in der weiteren Analyse diskutieren.

Unter der Annahme, dass jede Währung mit einem Vektor von Charakteristika  $x_{i,t}$  assoziiert werden kann und diese Charakteristika implizit alle nötigen Informationen zur Evaluierung der bedingten Wahrscheinlichkeitsverteilung der spezifischen Währungsrenditen beinhalten, können die optimalen Portfoliogewichte als eine Funktion der Währungscharakteristika formuliert werden:

$$w_{i,t} = f(x_{i,t}; \theta). \quad (5)$$

Wie in Barroso und Santa-Clara (2015) wird im Weiteren die folgende lineare Funktion der Portfoliogewichte verwendet:

$$w_{i,t} = \frac{\theta^T \hat{x}_{i,t}}{N_t}. \quad (6)$$

Dabei deklariert  $\theta$  einen Vektor mit Koeffizienten, welchen es mittels historischer Daten zu schätzen gilt. Die Charakteristika innerhalb des Vektors  $\hat{x}_{i,t}$  werden querschnittsweise standardisiert, sodass sie über alle Währungen zum Zeitpunkt  $t$  einen Durchschnittswert von 0 und eine Standardabweichung von 1 besitzen. Dadurch wird zum einen sichergestellt, dass die Querschnittsverteilung von  $\hat{x}_{i,t}$  stationär ist; bei Verwendung der nicht-standardisierten Charakteristika kann davon nicht bedenkenlos ausgegangen werden. Zum anderen wird somit garantiert, dass sich die optimalen Gewichte des Forward-Portfolios auf 0 aufsummieren.

Wie aus den Formeln (5) und (6) nachvollziehbar, ist der Koeffizienten-Vektor  $\theta$  währungs- und zeitunabhängig. Aus der Währungsunabhängigkeit folgt, dass das Gewicht einer Währung nur von deren Charakteristika abhängt und nicht etwa von deren historischer Renditen (Brandt et al., 2009). Demnach bekommen Währungen, welche sich in der Ausprägung ihrer Charakteristika gleichen, ein ähnliches Gewicht im Portfolio, unabhängig davon wie sich die historischen Renditen beider entwickelt haben. Die Zeitunabhängigkeit impliziert, dass der den bedingten erwarteten Nutzen maximierende Koeffizienten-Vektor zu jedem Zeitpunkt derselbe ist, weswegen er auch den unbedingten erwarteten Nutzen zu jedem Zeitpunkt maximiert (Brandt et al., 2009).

Gegeben der Annahmen von Währungs- und Zeitunabhängigkeit des Koeffizienten-Vektors kann das bedingte Optimierungsproblem abhängig von den Gewichten  $w_{i,t}$  aus Formel (2) zu einem unbedingten Optimierungsproblem abhängig vom Koeffizienten-Vektor  $\theta$  umformuliert werden:

$$\max_{\theta} E[u(r_{p,t+1})] = E[u(rf_{US,t:t+1} + \sum_{i=1}^{N_t} f(x_{i,t}; \theta) r_{i,t+1})]. \quad (7)$$

Innerhalb eines Datensatzes können somit die Koeffizienten anhand der folgenden Optimierung<sup>6</sup> geschätzt werden:

$$\max_{\theta} \frac{1}{T} \sum_{t=0}^{T-1} u(r_{p,t+1}) = \frac{1}{T} \sum_{t=0}^{T-1} u(rf_{US,t:t+1} + \sum_{i=1}^{N_t} (\frac{\theta^T \hat{x}_{i,t}}{N_t}) r_{i,t+1}). \quad (8)$$

Angelehnt an Brandt et al. (2009) und Barroso und Santa-Clara (2015) wird eine CRRA Nutzenfunktion unterstellt,

<sup>6</sup>Für Optimierungszwecke wurde auf das R-Package fBasics zurückgegriffen, während ich für die Berechnung der p-Werte über ein Bootstrapping-Verfahren das Package boot verwendet habe.

wobei  $\gamma$  den Grad der relativen Risikoaversion bezeichnet:

$$u(r_{p,t+1}) = \frac{(1 + r_{p,t+1})^{1-\gamma}}{1-\gamma} \quad (9)$$

Der wesentliche Vorteil im Vergleich zu einer Mittelwert-Varianz-Nutzenfunktion ist laut Barroso und Santa-Clara (2015), dass eine CRRA-Nutzenfunktion auch höhere Momente der Renditeverteilungen wie Schiefe und Kurtosis miteinbezieht. Damit werden innerhalb des Optimierungsverfahrens bspw. auch jene Charakteristika berücksichtigt, welche beim Senken des Crash-Risikos helfen, selbst wenn sie nicht zur Verbesserung der Sharpe Ratio beitragen. In der weiteren Analyse wird analog zu Barroso und Santa-Clara (2015)  $\gamma=4$  gewählt.<sup>7</sup>

Abschließend soll betont werden, dass die Optimierung aus Formel (8) die CRRA-Nutzenfunktion aus Formel (9) optimiert und nicht etwa den Abstand zwischen den durch die Charakteristika vorhergesagten und tatsächlichen Renditen. So ist es gemäß Barroso und Santa-Clara (2015) etwa möglich, dass eine Charakteristik durchaus einen Mehrwert zur Prognose künftiger Renditen liefern kann. Sollten diese Renditen allerdings gegeben der Nutzenfunktion des Investors ein unattraktives Rendite-Risiko-Verhältnis aufweisen, wird die Charakteristik im Zuge des Verfahrens dennoch als irrelevant klassifiziert.

### 2.3. Performance-Benchmarks

Als primäre und titelgebende Benchmark wird der diversifizierte Carry-Trade fungieren. Die Strategie besteht demnach aus einer 100%igen risikolosen Geldanlage in den USA und einem Long-Short-Portfolio aus Forward-Kontrakten. Die Rendite  $r_{fd,t+1}$  dieses Portfolios berechnet sich in jeder Periode wie folgt:

$$r_{fd,t+1} = rf_{US,t:t+1} + \sum_{i=1}^{N_t} w_{i,t}^{fd} r_{i,t+1}. \quad (10)$$

Hierbei entspricht das Portfoliogewicht  $w_{i,t}^{fd}$  der innerhalb des Optimierungsverfahrens verwendeten Charakteristik  $fd_{i,t}$ , die auf der prozentuellen Forward-Basis beruht und in Abschnitt 3.3.1 im Detail definiert wird. Da  $fd_{i,t}$  querschnittsweise standardisiert wurde, ist  $\sum_{i=1}^{N_t} w_{i,t}^{fd} \approx 0$ . Die Währungen im Terminkontraktportfolio streuen sich somit zu jedem Zeitpunkt typischerweise über das gesamte Währungsuniversum, wobei nicht das Vorzeichen der Forward-Basis einer Währung für die Portfoliogewichtung relevant ist, sondern der betragsmäßige Unterschied zwischen den Währungen im Querschnittsvergleich.

<sup>7</sup>Barroso und Santa-Clara (2015) treffen ihre Entscheidung hinsichtlich des Parameterwerts mit Verweis auf Bliss und Panigirtzoglou (2004), welche  $\gamma$  anhand der impliziten Risikoaversion schätzen, welche sie aus empirischen Daten zu einmonatigen Optionen auf Aktienindizes ableiten. Da gemäß Barroso und Santa-Clara (2015) die relevantesten Performancemaßzahlen (Sharpe Ratio, Schiefe, Kurtosis) skalenunabhängig sind, spielt die spezifische Wahl des Parameters keine entscheidende Rolle.

Wie DeMiguel et al. (2007) ausführen, bietet sich der  $\frac{1}{n}$ -Ansatz aufgrund seiner einfachen und verständlichen Berechnung sowie der üblicherweise geringen Implementierungskosten als ein natürlicher Maßstab für die Evaluierung komplexerer Strategien an. Dass die naive Diversifikation eine durchaus herausfordernde Benchmark sein kann, wurde in Abschnitt 2.1 bereits ausgeführt. In der vorliegenden Arbeit berechnet sich die Rendite dieser Benchmark in jeder Periode wie folgt:

$$r_{naive,t+1} = rf_{US,t:t+1} + \sum_{i=1}^{N_t} w_{i,t} r_{i,t+1}. \quad (11)$$

Das Gewicht  $w_{i,t}$  jeder Währung entspricht dabei  $1/N_t$ . Somit besteht das Forward-Portfolio zu gleichen Teilen aus Long-Positionen in allen zum Zeitpunkt  $t$  vorhandenen Währungen und einer aggregierten USD-Short-Position. Daraus folgt, dass gemäß der vorgestellten Definition des Leverage aus Formel (3) zu jedem Zeitpunkt eine Hebelwirkung von eins vorliegt.

Als letzte Benchmark und Vertreter einer anderen Asset-Klasse wird die Performance des amerikanischen Aktienmarkts herangezogen. Die nötigen Daten hierfür stammen von Kenneth Frenchs Internetseite<sup>8</sup>, welcher als Ursprung auf die CRSP-Datenbank verweist. Die Benchmark stellt gemäß der Internetseite einen wertgewichteten Index aus allen amerikanischen Firmen dar, die an der NYSE, AMEX und NASDAQ gelistet sind.

## 3. Selektion der Charakteristika

### 3.1. Übersicht der betrachteten Charakteristika

Wie in Abschnitt 2.1 bereits erwähnt, erfolgt die Selektion der Charakteristika für das vorgestellte Verfahren unter Zuhilfenahme der Gedanken und empirischen Ergebnisse anderer wissenschaftlicher Arbeiten, welche sich mit den Determinanten von Währungsrenditen beschäftigen. Als Ergebnis des Auswahlprozesses finden die folgenden Charakteristika Einzug in die weitere Analyse:

- Prozentuelle Forward-Basis bzw. Zinsunterschied (Carry)
- FX-Momentum
- Langfristige Trendumkehr (Realer Wechselkurs)
- Leistungsbilanzsaldo in Prozent des Bruttoinlandsprodukts (BIP)
- Produktionslücke in Prozent des potentiellen Bruttoinlandsprodukts (BIP)
- FX-Volatilität

<sup>8</sup>Zu finden unter folgendem Link: [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

Dabei war die Forschung von Barroso und Santa-Clara (2015), welche den Ansatz bereits im Devisenkontext angewendet haben, die Grundlage für die Auswahl. So habe ich die Charakteristika Carry, FX-Momentum, langfristige Trendumkehr und Leistungsbilanzsaldo dieser Arbeit entnommen und kann daher meine Ergebnisse mit denen der Autoren vergleichen. Mit der Produktionslücke und FX-Volatilität wurden darüber hinaus zwei Charakteristika inkludiert, die keine Verwendung in Barroso und Santa-Clara (2015) fanden, aber in jüngster Zeit von anderen Autoren in der wissenschaftlichen Literatur mit Aufmerksamkeit bedacht wurden. Als weitere mögliche Faktoren für zukünftige Arbeiten könnten z.B. wie in Nucera (2017) die Arbeitslosenrate in Betracht gezogen werden, oder ein Maß für Liquidität, wie es bspw. in der Arbeit von Mancini et al. (2013) verwendet wird. Um den Umfang der vorliegenden Arbeit aber nicht zu sprengen, liegt der Fokus auf den sechs genannten Faktoren.

Bevor nun der Literaturüberblick zu den selektierten Charakteristika beginnt, sollte zuletzt ein warnendes Wort von Cochrane (2017) angebracht werden. Dieser macht im Zusammenhang mit Aktien auf die Schwierigkeiten aufmerksam, die entscheidende(n) Zustandsvariable(n) in einem Asset-Pricing-Modell zur Erklärung von Überrenditen zu identifizieren. Sind vorgeschlagene Faktoren etwa korreliert, stellt sich die Frage, ob sie alle Schätzer für eine gemeinsam zugrundeliegende Zustandsvariable sind, oder ob das Modell weitere Zustandsvariablen beinhalten sollte. Als Folge betont Cochrane (2017) die Relevanz von ökonomischen Modellen zur Rationalisierung der aus den Faktormodellen abgeleiteten Risikoprämien – vor allem bei einer immer weiter steigenden Anzahl von in der Literatur vorgeschlagenen Faktoren.

### 3.2. Literaturüberblick

#### 3.2.1. FX-Momentum

Die Profitabilität von Momentum-Strategien in Devisenmärkten ist in der Literatur vielfach dokumentiert. Allerdings werden Arbeiten, welche Momentum im Querschnitt der Währungen betrachten, von der größeren Anzahl solcher, welche die Strategie innerhalb der Zeitreihe einer einzelnen Währung untersuchen<sup>9</sup>, abgegrenzt. Studien der letzteren Art werden im Folgenden demnach ausgeklammert, da die für den empirischen Teil gesuchten Charakteristika möglichst Aufschluss über die Wertentwicklung aller betrachteten Währungen geben sollen, um so zu einer optimalen Allokationsentscheidung zu kommen.

Okunev und White (2003) führen für die Währungen von acht entwickelten Volkswirtschaften eine Querschnittsbetrachtung innerhalb der Beobachtungsperiode von 1980 bis 2000 durch und analysieren in diesem Kontext Momentum-Strategien basierend auf unterschiedlich ausgestalteten gleitenden Durchschnittsverfahren. Sie präsentieren Evidenz für eine Outperformance von knapp 6% p.a., unabhängig von

der gewählten Heimatwährung und der im Detail gewählten Methode des gleitenden Durchschnittsverfahrens.

Die Arbeit von Menkhoff et al. (2012b) betrachtet die Daten ebenfalls im Querschnitt, allerdings ist der Datensatz im Vergleich zur oben genannten Arbeit mit 48 Währungen und einer Beobachtungsperiode von 1976 bis 2010 durchaus extensiver. Neben der Konstatierung einer signifikanten Outperformance der Momentum-Strategie gehört zu ihrem wesentlichen Ergebnis auch die Feststellung, dass traditionelle Benchmark-Faktoren wie etwa der Carry-Trade oder Volatilitätsrisiko jene Überrenditen nicht erklären können, wie zuvor schon Burnside et al. (2011) empirisch darlegten. Zusätzlich weisen Menkhoff et al. (2012b) auf eine geringe Korrelation der Renditen zwischen diesen Strategien hin. Des Weiteren machen sie darauf aufmerksam, dass die Momentum-Strategie zwar durchaus sensitiv bzgl. mit ihr einhergehenden Transaktionskosten ist, diese aber die Outperformance nicht vollständig nivellieren. Zudem identifizieren die Autoren Muster in den Daten, welche sie als konsistent mit verhaltensökonomischen Erklärungen zu Über- und Unterreaktionen von Investoren betrachten.

Asness et al. (2013) untersuchen Value- und Momentum-Strategien über diverse Assetklassen hinweg. Sie stellen für Momentum eine Korrelation der Renditen über die untersuchten Assetklassen hinweg fest und konkludieren auf einen globalen Risikofaktor im Zusammenhang mit Momentum. Zudem konstatieren sie eine negative Korrelation zwischen Value- und Momentum-Strategien innerhalb und über alle untersuchten Assetklassen hinweg, was Diversifikationsvorteile verspricht.

Zuletzt sei erneut auf Barroso und Santa-Clara (2015) verwiesen, welche Momentum ebenfalls als Charakteristik in ihrer Portfoliooptimierung inkludieren, wobei sie den Faktor als kumulative Währungsaufwertung der letzten drei Monate definieren. Innerhalb ihres In-Sample-Tests kann für Momentum sowohl bei einer gesamtheitlichen Betrachtung aller Charakteristika in der Optimierung als auch isoliert betrachtet Signifikanz nachgewiesen werden. Dieses Resultat können sie anschließend im Zuge ihres Out-of-Sample-Tests bestätigen. Ebenfalls von den Autoren als Robustheitscheck thematisiert werden Transaktionskosten, welche im Zusammenhang mit einer Momentum-Strategie substantiell sein können. Out-of-Sample und unter Berücksichtigung der zuvor im In-Sample-Test als signifikant identifizierten Faktoren verliert Momentum seine Profitabilität. Auf Basis der gleichen Betrachtung unter Einbezug aller von ihnen selektierten Charakteristika stellen die Autoren allerdings weiterhin eine wesentliche Outperformance des optimalen Portfolios fest, weswegen sie nahelegen, dass nicht wie oft üblich die Transaktionskosten einer individuellen Strategie, sondern mehrerer Strategien in Interaktion betrachtet werden sollten.

In Summe ist die Renditeprädiktabilität von Momentum in der Literatur ausführlich dokumentiert. Zur Beantwortung der Frage, warum sie besteht, gibt es allerdings kein einheitliches Credo. So existieren etwa im Aktiensegment auf Jegadeesh und Titman (1993) zurückgehende Erklärungsansätze, welche die mit Momentum assoziierten Überrenditen in

<sup>9</sup>Einen Literaturüberblick bzgl. dieser oft „Technical Trading Rules“ genannten Ansätze liefern bspw. Menkhoff und Taylor (2007).



einem verhaltensökonomischen Rahmen als das Resultat von Über- bzw. Unterreaktionen von Investoren betrachten.<sup>10</sup> Zudem wurden verschiedene Versuche unternommen, die Profitabilität der Momentum-Strategie mit einhergehendem Risiko in Verbindung zu bringen – dies gelang jedoch nicht immer mit Erfolg. Liu und Zhang (2008) listen einige Arbeiten auf, die sich zum Ziel setzten, Momentum im Aktiensegment mittels bekannter Risikofaktoren wie bspw. Marktrisiko oder der drei Fama-French-Faktoren zu erklären und daran scheiterten. Sie führen aber auch Arbeiten an, welche Evidenz für Erklärungen im Zusammenhang mit Liquiditäts- bzw. Konsumrisiko konstatieren.

### 3.2.2. Realer Wechselkurs

Einem Gedanken von Shiller (1981) folgend argumentiert Cochrane (2007), welcher den Zusammenhang von künftigen Aktienrenditen und der heutigen Dividendenrendite untersucht, dass gegeben einem Barwertmodell sowie einer stationären Dividendenrendite die Schwankungen der Dividendenrendite eine Vorhersagekraft hinsichtlich der zukünftigen Aktienrendite oder dem zukünftigen Dividendenwachstum oder beidem innehaben muss. Ausgehend von der Nullhypothese eines effizienten Marktes, welcher keine Renditevorhersagbarkeit zulässt und die Schwankung in der Dividendenrendite dem künftigen Dividendenwachstum zuschreibt, kommt er zum Schluss, dass bereits die fehlende Prognostizierbarkeit von Dividendenwachstum in den Daten Evidenz zum Verwerfen der Nullhypothese liefert. Cochrane (2011) weist ebenfalls darauf hin, dass die empirische Aussagekraft der Dividendenrendite bzgl. der künftigen Preisentwicklung steigt, je weiter die vorhergesagten Renditen in der Zukunft liegen. Van Binsbergen und Kojien (2010) stellen in diesem Zusammenhang fest, dass die erwartete Aktienrendite persistenter als die erwartete Dividendenwachstumsrate ist.

Verschiedene Arbeiten, wie z.B. Froot und Ramadorai (2005) oder Engel und West (2010), haben diesen Gedanken des Barwertansatzes bereits auf Deviseninvestments übertragen. Aktuelle Studien von Dahlquist und Penasse (2017) sowie Balduzzi und Chiang (2017) betrachten in diesem Kontext neben dem bekannten Einfluss des Zinsunterschieds, welcher wie eingangs dargelegt mit Abweichungen von der ungedeckten Zinsparität zusammenhängt, die Auswirkung des realen Wechselkurses auf die zukünftigen Devisenrenditen. Beim realen Wechselkurs handelt es sich um die nominale Größe nach Inflationsbereinigung mittels in- und ausländischem Preisindex (Burda und Wyplosz, 2009). Die Autoren unterstellen hierbei auf Basis einer Arbitrageüberlegung, dass langfristig die Kaufkraftparität Bestand hat. Ähnlich wie im geschilderten Fall für Aktienrenditen erlaubt ihnen dies, im Kontext eines Barwertmodells und gegeben der Annahme eines stationären realen Wechselkurses<sup>11</sup>, den

heutigen realen Wechselkurs hinsichtlich seiner Aussagekraft bzgl. des zukünftigen Zinsunterschieds und der zukünftigen Währungsrendite aufzuspalten.

Dahlquist und Penasse (2017) stellen empirisch fest, dass der Hauptprädiktor von Währungsrenditen über einen längeren Zeithorizont mit Abweichungen von der Kaufkraftparität zusammenhängt, während Konvergenzen von der ungedeckten Zinsparität die Renditen über einen kurzfristigen Zeitraum beeinflussen. Sie formulieren darauf aufbauend die vergleichende Theorie, dass persistente Risikoprämien die langfristigen Abweichungen von der Kaufkraftparität im Devisenmarkt, wie auch die langfristigen Bewegungen der Dividendenrendite im Aktienkontext, erklären. Ähnlich hierzu lehnen auch Balduzzi und Chiang (2017) die Nullhypothese ab, welche keine Vorhersagbarkeit der zukünftigen Währungsrenditen durch den effektiven Wechselkurs unterstellt.

Ähnlich zu ihrem im vorigen Unterkapitel beschriebenen Ergebnis in Bezug auf Momentum-Strategien, stellen Asness et al. (2013) ebenfalls für Value-Strategien eine Parallelbewegung der Renditen über die untersuchten Assetklassen hinweg fest, was die Autoren analog als Indikation für einen globalen Risikofaktor im Zusammenhang mit „Value“ sehen. Sie definieren Value im Währungsbereich auf Basis des realen Wechselkurses und der Kaufkraftparität.

Weiterhin machen Jordà und Taylor (2012) auf die Diversifikationsvorteile sowie eine verbesserte Crash-Resistenz in der letzten globalen Finanzkrise aufmerksam, welche durch die Berücksichtigung eines Signals basierend auf dem realen Wechselkurs im Kontext der Carry-Trade-Strategie entstehen. Das verwendete Signal leitet sich dabei konzeptuell ähnlich zu bereits genannten Studien aus einer langfristigen Kaufkraftparitätsüberlegung ab.

Zuletzt sei auf Barroso und Santa-Clara (2015) verwiesen, welche in ihrer Arbeit den realen Wechselkurs ebenfalls als Charakteristik innerhalb des Ansatzes zur Portfoliooptimierung von Brandt et al. (2009) nutzen. Die besagte Größe fließt in zwei ihrer Charakteristika ein, welche sie mit Realer Wechselkurs („Real Exchange Rate“) und Langfristige Trendumkehrung („Long-Term Reversal“) bezeichnen. Im Zuge ihres In-Sample-Tests und unter Berücksichtigung aller von ihnen selektierten Faktoren stellen sie fest, dass von beiden Charakteristika nur die langfristige Trendumkehrung statistische Signifikanz aufweist; isoliert betrachtet wird sie jedoch auch verworfen. In einem weiteren Out-of-Sample-Test, basierend auf den zuvor im In-Sample-Test identifizierten signifikanten Größen, inkludieren sie daher nur jenen statistisch signifikanten Faktor. Ihr festgestelltes Ergebnis trifft in Folge weiterhin zu.

### 3.2.3. Leistungsbilanzsaldo

Die Leistungsbilanz als Teil der Zahlungsbilanz einer Volkswirtschaft definiert sich als Summe aus Handels- und

<sup>10</sup>Bspw. thematisieren die Arbeiten von Daniel et al. (1998), Barberis et al. (1998) sowie Hong und Stein (1999) verschiedene psychologische Verhaltensmuster als Ursachen.

<sup>11</sup>Verschiedene Arbeiten von bspw. Rogoff (1996), Lothian und Taylor

(1996), Cheung et al. (2005) oder Burstein und Gopinath (2014) thematisieren die Stationarität des realen Wechselkurses bzw. testen auf den Bestand der Kaufkraftparität und kommen zu gemischten Ergebnissen. Lustig et al. (2017) argumentierten jüngst gar für eine Stationarität nominaler Wechselkurse.



Dienstleistungsbilanz, Bilanz der Erwerbs- und Vermögenseinkommen sowie der Übertragungsbilanz, siehe hierzu etwa Burda und Wyplosz (2009). Sie bzw. ihr Saldo wird bereits seit Jahrzehnten von Forschern als mögliche relevante Determinante von Wechselkursbewegungen in Betracht gezogen und diskutiert. So existieren bereits aus den 1970er und 1980er Jahren wissenschaftliche Arbeiten, die neben den klassischen monetären Wechselkursmodellen, welche üblicherweise u.a. auf Annahmen hinsichtlich des Bestands der Kaufkraftparität sowie der ungedeckten Zinsparität basieren, sogenannte „Portfolio-Balance-Modelle“ konstruieren<sup>12</sup>.

In diesen gedanklich auf Tobin (1969) zurückgehenden Modellen bestimmt sich der Wechselkurs nicht ausschließlich aus Geldangebot und -nachfrage wie in den monetären Modellen. Vielmehr ist das Angebot von und die Nachfrage nach allen einheimischen und ausländischen Assets für den Devisenkurs relevant, wobei einheimische und ausländische Assets keine perfekten Substitute sind. Diesem Gedanken folgend stellt der Leistungsbilanzsaldo einer kleinen Volkswirtschaft<sup>13</sup> in einem solchen Modell eine Veränderung im Angebot von ausländischen Assets dar. Ein Überschuss erhöht somit das Angebot von ausländischen Assets und reduziert deren Preis, was wiederum der Logik des Anpassungsmechanismus dieser Gleichgewichtsmodelle folgend zu einer Aufwertung der einheimischen Währung führen sollte (Frankel und Rose, 1995). In einer aktuelleren Arbeit entwickeln etwa Gabaix und Maggiori (2015) diesen Ansatz weiter und stellen u.a. ein Modell vor, in welchem der Wechselkurs zugleich von globalen Handelsungleichgewichten zwischen den Ländern sowie dem Willen zur Risikoübernahme der agierenden Investoren abhängt. Letztere sind demnach aufgrund ihrer gegebenen Risikoübernahmefähigkeit nur bis zu einem gewissen Grad dazu geneigt Währungsungleichgewichte gegen eine adäquate Risikoprämie zu absorbieren.

Die Vorhersagbarkeit von Wechselkursbewegungen anhand beobachtbarer makroökonomischer Fundamentaldaten basierend auf der theoretischen Fundierung von monetären Modellen oder Portfolio-Balance-Modellen wurde in verschiedenen empirischen Arbeiten untersucht. Meese und Rogoff (1983) konstatieren, dass für die untersuchten Währungen bei einem Betrachtungshorizont von einem bis zwölf Monaten ein Random-Walk-Modell ebenso gut wie jedes der getesteten Modelle abschneidet. Frankel und Rose (1995) bekräftigen den Fund von Meese und Rogoff (1983) und formulieren in ihrem Literaturüberblick den zusammenfassenden Konsens von relativ geringer empirischer Erklärungskraft der Modellansätze. Auch Lee und Chinn (2006) bestätigen dieses Ergebnis und vernetzen es gedanklich mit der zen-

tralen Erkenntnis ihrer Arbeit, welcher zur Folge temporäre Schocks relevanter für die Entwicklung der Leistungsbilanz sind als permanente Schocks, welche hauptsächlich auf die Wechselkursbewegung wirken.

Im Gegensatz zu den bisher erwähnten Arbeiten schließen Gourinchas und Rey (2007), welche ein der Leistungsbilanz sehr ähnliches Signal verwenden, auf eine positive Erklärungskraft hinsichtlich zukünftiger Währungsrenditen In- sowie Out-of-Sample. Ebenso präsentieren Della Corte et al. (2016) Evidenz für die Existenz einer Risikoprämie, welche Investoren für das Halten einer Währung von Nettokreditnehmerländern fordern, da solche Währungen dazu neigen in schlechten Zeiten abzuwerten. Die Autoren weisen nicht nur nach, dass der von ihnen konstruierte globale Risikofaktor die Querschnittsvariation von Währungsüberrenditen, sondern auch den von Renditen anderer relevanter Asset-Klassen wie Aktien, Bonds und Rohstoffe erklärt.

Barroso und Santa-Clara (2015) verwenden ebenfalls die Leistungsbilanz als renditenbeschreibenden Faktor innerhalb des Optimierungsansatzes von Brandt et al. (2009). Nachdem die Autoren aber bereits im In-Sample-Test keine statistische Signifikanz feststellen können, exkludieren sie den Faktor in ihrem weiteren Out-of-Sample-Test.

### 3.2.4. Produktionslücke

Wie im vorherigen Unterkapitel zum Leistungsbilanzsaldo dargestellt ist der Erfolg der empirischen Forschung begrenzt, das Risiko von Währungen anhand von makroökonomischen Fundamentaldaten zu erklären, obwohl die ökonomische Intuition zumindest auf einen langfristigen Zusammenhang zwischen beidem schließen lässt. Riddiough und Sarno (2017) ziehen in diesem Zusammenhang den Vergleich zum Risiko eines Unternehmens, welches der Theorie nach von dessen Fundamentaldaten abhängen sollte. Basierend auf der Arbeit von Verdelhan (2010) konstruieren sie ein theoretisches Modell, welches Währungsüberrenditen mit den Konjunkturzyklen der verschiedenen Länder in Verbindung bringt. Der einfachen Intuition dieses Modells folgend sind Investoren aus Ländern mit schwacher Wirtschaft risikoaverser als solche aus Ländern mit starker Wirtschaft und verlangen daher eine Risikoprämie für das Halten ausländischer Währungen.

Die aus dem Modell ableitbare Hypothese, welche es zu testen gilt, bezieht sich auf die Produktionslücke<sup>14</sup> und hängt nicht von Daten zu aggregiertem Länderkonsum ab, welche schwer verfügbar bzw. oftmals unzureichend erhobenen sind. Der Hypothese folgend sollte eine Strategie, welche Währungen von Ländern mit hoher Produktionslücke kauft und solche von Ländern mit geringer Produktionslücke verkauft, zu Outperformance führen. Riddiough und Sarno (2017) finden Evidenz für diese Aussage und deuten darauf hin, dass eine geringe Korrelation zwischen ihrer und anderen Währungsinvestmentstrategien aus der Literatur besteht, wie et-

<sup>12</sup>Hier kann bspw. auf Kouri (1976), Dornbusch und Fischer (1980) oder Hooper und Morton (1982) verwiesen werden. Für eine vergleichende Gegenüberstellung von monetären Modellen und Portfolio-Balance-Modellen siehe z.B. Frankel und Rose (1995).

<sup>13</sup>Beim vereinfachten Modell einer kleinen Volkswirtschaft sind internationale Kapitalströme annahmegemäß ausschließlich in ausländischer Währung denominated, während bei realistischeren Modellen das Portfolioverhalten sowohl der einheimischen wie auch ausländischen Bevölkerung für das Marktgleichgewicht relevant ist, siehe hierzu etwa Frankel und Rose (1995).

<sup>14</sup>Die Produktionslücke bezeichnet die prozentuale Abweichung des Bruttoinlandsprodukts von seinem langfristig nachhaltigen Trend (Burda und Wyplosz, 2009).

wa dem Carry-Trade oder dem von Menkhoff et al. (2012a) vorgeschlagenen Volatilitätsfaktor<sup>15</sup>. In Folge dessen attestieren Riddiough und Sarno (2017) ihrer Strategie nicht nur reizvolle risikoadjustierte Renditen, sondern auch vorteilhafte Diversifikationseigenschaften.

### 3.2.5. FX-Volatilität

Nachdem Volatilität als möglicher Risikofaktor bereits im Aktiensegment zur Erklärung von Querschnittsrenditen herangezogen wurde<sup>16</sup>, gibt es auch Arbeiten, die Evidenz für die Vorhersage von Devisenrenditen konstatieren.

Zu nennen ist hierbei Menkhoff et al. (2012a), welche betrachten, inwiefern die Überrenditen des Carry-Trades teilweise als Kompensation für ein FX-Volatilitätsrisiko erklärt werden können. Diese Hypothese leiten sie aus dem ökonomischen Gedanken ab, welchem zufolge etwa eine erhöhte unerwartete Volatilität einen negativen Effekt auf das Risiko-Rendite-Verhältnis eines Investors hat und demnach einem unvorteilhaften Umweltzustand gleichkommt. Investoren verlangen also eine zusätzliche Kompensation für Währungen, die bei hoher unerwarteter Volatilität schlecht performen, während sie bereit sind ein Premium für Währungen aufzubringen, die im gleichen Umweltzustand profitabel sind.

Zur Durchführung ihres empirischen Tests definieren sie eine für die globale FX-Volatilität stehende Proxy-Variable basierend auf Beobachtungen zur täglichen realisierten Volatilität des vorangegangenen Monats und untersuchen die zeitliche Änderung in dieser Proxy-Variable als nichtgehandelten Risikofaktor, ähnlich wie es Ang et al. (2006) mit ihrem definierten Volatilitätsfaktor im Aktiensegment handhaben. Sie stellen fest, dass die Renditen von Hochzinswährungsportfolios einen negativen Zusammenhang mit Änderungen in der globalen FX-Volatilität aufweisen, während Niedrigzinswährungsportfolios bei Zunahme des Risikofaktors positive Renditen und somit eine Art Hedge liefern. Der eingangs vorgestellte Carry-Trade performt demnach tendenziell schlechter in Zeiten von Marktunruhen und -chaos. Davon ausgehend argumentieren die Autoren für eine Erklärung der Renditen des Carry-Trades als Kompensation für zeitvariierendes Risiko. Zusätzlich konstatieren Menkhoff et al. (2012a), dass der definierte Risikofaktor nicht nur mit anderen Proxy-Variablen für Marktliquidität (bspw. Bid-Ask-Spreads oder TED-Spread) korreliert ist, sondern diese tatsächlich im direkten Vergleich in Sachen Erklärungskraft bzgl. der Renditen des Carry-Trades dominiert. Sie weisen hierbei Brunnermeier et al. (2008) folgend auf den engen Zusammenhang zwischen den Konzepten von beobachtbarer Marktvolatilität und nicht beobachtbarer Marktliquidität bei der Erklärung von Devisenrenditen hin. Ebenfalls weisen sie nach, dass mit Hilfe des Faktors auch die Querschnittsrenditen anderer Strategien<sup>17</sup> wie z.B. etwa Momentum in Devisenmärkten rationalisiert werden können.

In einer anderen Studie untersuchen Corte et al. (2016) ebenso die empirische Aussagekraft eines Volatilitätsfaktors als Prädiktor für künftige Währungsrenditen, ohne dabei im Detail ein formales theoretisches Modell für dessen Erklärung zu liefern. Anders als bei Menkhoff et al. (2012a) konstruieren die Autoren den besagten Faktor auf der grundlegenden Idee eines Volatilitätsswap-Kontrakts, welcher dem Kontraktkäufer eine Auszahlung in Abhängigkeit der realisierten Volatilität eines Basisinstruments über die Laufzeit des Vertrags zusichert, für die er als Gegenleistung einen vorher festgelegten Preis („Swap Rate“) zahlen muss. Da der Kontraktwert über eine Barwertbetrachtung der erwarteten zukünftigen Auszahlung in einer risikoneutralen Welt ermittelt wird, sollte die Swap Rate einer risikoneutralen Erwartung der besagten realisierten Volatilität gleichen. Die Autoren synthetisieren diese Swap Rate mittels eines Ansatzes von Britten-Jones und Neuberger (2000) aus Preisen von Call-Optionen mit Ablaufdatum zum Zeitpunkt des Auslaufs des hypothetischen Volatilitätsswap-Kontrakts. Darauf aufbauend konstruieren sie die Volatilitätsrisikoprämie als Differenz aus der ex-post realisierten Volatilität, gemessen über eine Vorperiode, abzüglich der ex-ante risikoneutralen Erwartung der künftig realisierten Volatilität über die Betrachtungsperiode. Währungen mit einer relativ hohen (niedrigen) Faktorgewichtung gelten demnach als günstig (teuer), da sie über die Vorperiode eine hohe (geringe) realisierte Volatilität ausweisen, aber nach wie vor günstig (teuer) über Optionen zu versichern sind. Das physikalische Wahrscheinlichkeitsmaß liegt demnach also höher (niedriger) als das risikoneutrale Wahrscheinlichkeitsmaß.

Ähnlich der zitierten Arbeit von Menkhoff et al. (2012a) konstruieren auch Corte et al. (2016) Währungsportfolios, sortiert nach Faktoren wie ihrem soeben präsentierten Volatilitätsfaktor sowie traditionellen Ansätzen wie etwa Carry, Momentum oder Value<sup>18</sup>. Sie finden Evidenz für eine deutliche Outperformance der Volatilitätsstrategie gegenüber den anderen häufig benutzten Faktoren, gemessen als Rendite per Einheit Volatilität. Zudem weisen die Autoren auf einen möglichen Diversifikationsvorteil der Strategie hin, da sie eine nur geringe Korrelation zu den anderen betrachteten Strategien feststellen. In diesem Zusammenhang betonen sie als wesentlichen Unterschied zu den traditionellen Ansätzen, dass die Performance der Volatilitätsstrategie maßgeblich auf der Vorhersagbarkeit von Währungskursen basiert, während etwa beim Carry-Trade der Zinsunterschied die wesentliche Komponente der Renditen darstellt und ein vergleichsweise geringerer Teil der Renditeprädiktabilität auf Variation in den zukünftigen Wechselkursen beruht. Schließlich zeigen die Autoren auf, dass auch andere verbreitete Asset-Pricing-Modelle außerhalb des Devisensegments, wie etwa die Faktorenmodelle von Fama und French (1993) oder Fung und Hsieh (2004), die Renditen der Volatilitätsstrategie nicht

<sup>15</sup>Mehr hierzu in Unterkapitel 3.2.5.

<sup>16</sup>Thematisiert wird dies bspw. von Ang et al. (2006) sowie Adrian und Rosenberg (2008).

<sup>17</sup>Bemerkenswerterweise attestieren die Autoren dem Faktor etwa auch

Aussagekraft bzgl. der Querschnittsrenditen amerikanischer Unternehmensanleihen oder einer Momentum-Strategie im US-Aktienmarkt.

<sup>18</sup>Die Autoren definieren Value ebenfalls anhand des realen Wechselkurses, siehe dazu vergleichend Unterkapitel 3.2.2.

rationalisieren können.

Zuletzt sei noch auf die Arbeiten von Bhansali (2007) sowie Clarida et al. (2009) verwiesen, welche ebenfalls einen Zusammenhang zwischen den Carry-Renditen und der Volatilität in Währungsmärkten feststellen. In diesem Kontext präsentieren sie u.a. Optionsstrategien, welche persistente Überrenditen generieren, deren Outperformance jedoch ähnlich zum Carry-Trade bei einem Anstieg der Volatilität versiegt. Des Weiteren konstatieren Clarida et al. (2009) etwa eine hohe Sensitivität der Ergebnisse jener in Abschnitt 1.1 erwähnten bekannten Regression zum Test auf Bestand der ungedeckten Zinsparität hinsichtlich des Volatilitätsregimes. Demnach liefert diese zu Zeiten geringer Volatilität einen signifikanten negativen Koeffizienten, während sich das Vorzeichen des Ergebnisses in Perioden hoher Volatilität dreht und weiterhin signifikant bleibt. Bemerkenswerterweise löst diese Feststellung allerdings keineswegs die Thematik einer Verletzung der ungedeckten Zinsparität, da der von den Autoren geschätzte Koeffizient größer 1 ist, d.h. dass Niedrigzinswährungen tatsächlich stärker aufwerten als durch die ungedeckte Zinsparität diktiert.

### 3.3. Methodologie der Charakteristika

#### 3.3.1. sProzentuelle Forward-Basis bzw. Zinsunterschied (Carry)

Gegeben der Annahme, dass die gedeckte Zinsparität hält, entspricht die prozentuale Forward-Basis

$$FB_{i,t:t+1} = \frac{F_{i,t:t+1} - S_{i,t}}{S_{i,t}} \quad (12)$$

der Währung  $i$  im Verhältnis zum USD gemessen zum Zeitpunkt  $t$  näherungsweise dem Zinsunterschied der zugrundeliegenden Länder (Riddiough und Sarno, 2017). Den Bestand der gedeckten Zinsparität stellen zwar bspw. Akram et al. (2008) fest, allerdings machen Du et al. (2017) in einer neueren Studie auf teils massive Abweichungen seit der globalen Finanzkrise im Jahr 2007 aufmerksam.

Die Charakteristik  $f d_{i,t}$  definiere ich wie in Barroso und Santa-Clara (2015) als Standardisierung der prozentuellen Forward-Basis  $FB_{i,t:t+1}$ :

$$f d_{i,t} = \frac{FB_{i,t:t+1} - \mu_{FB_t}}{\sigma_{FB_t}}, \quad (13)$$

wobei  $\mu_{FB_t}$  und  $\sigma_{FB_t}$  dabei Durchschnitt und Standardabweichung des Querschnitts aller zum Zeitpunkt  $t$  verfügbaren Währungen darstellen. Die Charakteristik misst also die Forward-Basis in Standardabweichungen über bzw. unter dem Mittelwert aller Währungen und hat über den Querschnitt aller Währungen einen Mittelwert von null. Demnach ist nicht das Vorzeichen der Forward-Basis einer Währung für die Charakteristik und daraus folgende Portfoliogewichtung relevant, sondern der betragsmäßige Unterschied zwischen den Währungen im Querschnittsvergleich. Außerdem impliziert die Definition, dass die Strategie neutral im Sinne

der Heimatwährung, dem USD, ist (Barroso und Santa-Clara, 2015).

#### 3.3.2. FX-Momentum

Für die Charakteristik  $mom_{i,t}$  wird wie in Kroencke et al. (2013) oder Barroso und Santa-Clara (2015) die kumulative Währungsaufwertung der letzten drei Monate verwendet, welche über den Querschnitt aller Währungen zum Zeitpunkt  $t$  standardisiert wird.

Zu jedem Zeitpunkt  $t$  wird für alle vorhandenen Währungen die kumulative Rendite der vorherigen drei Monate berechnet:

$$MOM_{i,t} = -\left(\frac{S_{i,t}}{S_{i,t-3}} - 1\right). \quad (14)$$

Ist  $MOM_{i,t}$  positiv (negativ), hat Währung  $i$  gegen den USD in den vorherigen drei Monaten aufgewertet (abgewertet). Anschließend kann  $mom_{i,t}$  gebildet werden, indem  $MOM_{i,t}$  wie folgt standardisiert wird:

$$mom_{i,t} = \frac{MOM_{i,t} - \mu_{MOM_t}}{\sigma_{MOM_t}}, \quad (15)$$

wobei  $\mu_{MOM_t}$  und  $\sigma_{MOM_t}$  dabei Durchschnitt und Standardabweichung des Querschnitts aller zum Zeitpunkt  $t$  verfügbaren Währungen darstellen. Die Charakteristik ist für jene Währungen positiv, welche in den letzten drei Monaten vergleichsweise stark gegen den USD aufgewertet haben.

#### 3.3.3. Langfristige Trendumkehr (Realer Wechselkurs)

Die Charakteristik  $rev_{i,t}$  basiert wie in Barroso und Santa-Clara (2015) auf der kumulativen Aufwertung des realen Wechselkurses, betrachtet über  $h$  Perioden und ausgedrückt als folgende Indexnummer<sup>19</sup>:

$$Q_{i,h,t} = \frac{S_{i,t} CPI_{i,h-2} CPI_{US,t-2}}{S_{i,h} CPI_{i,t-2} CPI_{US,h-2}}, \quad (16)$$

wobei  $CPI_{i,t}$  die Inflationsschätzung des mit der Währung  $i$  assoziierten Landes in Form des Konsumentenpreisindexes darstellt. Identisch zu Barroso und Santa-Clara (2015) wird  $h = t - 60$  gewählt, was einem Zeitraum von fünf Jahren entspricht.

Um zur Charakteristik  $rev_{i,t}$  zu gelangen, wird  $Q_{i,h,t}$  über den Querschnitt aller Währungen standardisiert:

$$rev_{i,t} = \frac{Q_{i,h,t} - \mu_{Q_{h,t}}}{\sigma_{Q_{h,t}}}, \quad (17)$$

wobei  $\mu_{Q_{h,t}}$  und  $\sigma_{Q_{h,t}}$  dabei Durchschnitt und Standardabweichung des Querschnitts aller zum Zeitpunkt  $t$  verfügbaren Währungen darstellen. Somit ist die Charakteristik wie

<sup>19</sup>Erläuterungen zur in der Formel ersichtlichen Zeitverzögerung von zwei Monaten in den CPI-Daten finden sich in Abschnitt 4.2.

bei Barroso und Santa-Clara (2015) für Währungen positiv, die in den letzten fünf Jahren im Schnitt stark gegen den US-Dollar abgewertet haben.

### 3.3.4. Leistungsbilanzsaldo in Prozent des Bruttoinlandsprodukts

Der Leistungsbilanzsaldo wird, wie bei Barroso und Santa-Clara (2015), im Zuge der Verwendung als Charakteristik  $ca_{i,t}$ , über den Querschnitt aller Währungen standardisiert:

$$ca_{i,t} = \frac{CA_{i,t} - \mu_{CA_t}}{\sigma_{CA_t}}. \quad (18)$$

$CA_{i,t}$  stellt dabei den Leistungsbilanzsaldo in Prozent des Bruttoinlandsprodukts der Währung  $i$  zum Zeitpunkt  $t$  dar, während  $\mu_{CA_t}$  und  $\sigma_{CA_t}$  den Durchschnitt und die Standardabweichung des Querschnitts aller zum Zeitpunkt  $t$  verfügbaren Währungen bezeichnen.

### 3.3.5. Produktionslücke in Prozent des potentiellen Bruttoinlandsprodukts

Ähnlich zum Vorgehen bei den von Barroso und Santa-Clara (2015) selektierten Charakteristika definiere ich auch  $opg_{i,t}$  als Standardisierung der Produktionslücke über den Querschnitt aller Währungen:

$$opg_{i,t} = \frac{OPG_{i,t} - \mu_{OPG_t}}{\sigma_{OPG_t}}. \quad (19)$$

$OPG_{i,t}$  bezeichnet dabei die Produktionslücke in Prozent des potentiellen Bruttoinlandsprodukts der Währung  $i$  zum Zeitpunkt  $t$ .  $\mu_{OPG_t}$  und  $\sigma_{OPG_t}$  sind Platzhalter für Durchschnitt und Standardabweichung des Querschnitts aller zum Zeitpunkt  $t$  verfügbaren Währungen.

### 3.3.6. FX-Volatilität

Im Zuge der Definition von Charakteristik  $vol_{i,t}$  wird zunächst für jede Währung  $i$  an jedem Tag  $\tau$  die tägliche Rendite berechnet:

$$r_{i,\tau} = \frac{S_{i,\tau}}{S_{i,\tau-1}} - 1. \quad (20)$$

Anschließend wird pro Währung  $i$  für jeden Monat  $t$ , welcher als Datenpunkt für die Optimierung fungieren soll, die Variable  $VOL_{i,t}$  bestimmt. Diese entspricht der Standardabweichung, welche auf Basis der täglichen Renditen seit der Vorperiode  $t-1$  geschätzt wurde.

Um zur Charakteristik  $vol_{i,t}$  zu gelangen, wird  $VOL_{i,t}$  über den Querschnitt aller Währungen standardisiert:

$$vol_{i,t} = \frac{VOL_{i,t} - \mu_{VOL_t}}{\sigma_{VOL_t}}. \quad (21)$$

Sie weist demnach einen positiven (negativen) Wert für jene Währungen auf, die im Schnitt eine vergleichsweise hohe (geringe) realisierte Volatilität über die Vorperiode verzeichneten.

## 4. Daten

### 4.1. Benötigte Daten und Vorstellung der Datenquellen

Für den gewählten Portfoliooptimierungsansatz und die dazugehörigen selektierten Charakteristika wurden folgende Datentypen als relevant identifiziert:

- Mid-Spot Rates der Länderwährungen
- Mid-1-Monat-Forward Rates der Länderwährungen
- Preisindex der Länder
- Leistungsbilanzsaldo in Prozent des BIPs der Länder
- Produktionslücke in Prozent des potentiellen BIPs der Länder
- Risikoloser Zinssatz für eine Haltedauer von einem Monat.

Um der Zielsetzung eines in Breite und Länge möglichst umfangreichen Datensatzes zur weiteren Analyse Rechnung zu tragen, wurden die aufgeführten Informationen als historische Zeitreihen mit monatlicher Frequenz für die folgenden 42 Länder (entsprechende Währungscode in Klammern) mit flexiblem Wechselkurs ab Januar 1974, ein Jahr nach Zusammenbrechen des Bretton-Woods-Systems, bis Ende Februar 2018 abgefragt:

- Ägypten (EGP)
- Australien (AUD)
- Belgien (BEF)
- Brasilien (BRL)
- Bulgarien (BGN)
- Chile (CLP)
- Dänemark (DKK)
- Deutschland (DEM)
- Eurozone (EUR)
- Finnland (FIM)
- Frankreich (FRF)
- Griechenland (GRD)
- Indien (INR)
- Indonesien (IDR)
- Irland (IEP)
- Island (ISK)
- Italien (ITL)
- Japan (JPY)
- Kanada (CAD)
- Kroatien (HRK)



- Malaysia (MYR)
- Mexiko (MXN)
- Neuseeland (NZD)
- Niederlande (NLG)
- Norwegen (NOK)
- Österreich (ATS)
- Philippinen (PHP)
- Polen (PLN)
- Portugal (PTE)
- Republik China /Taiwan (TWD)
- Russland (RUB)
- Schweden (SEK)
- Schweiz (CHF)
- Slowakei (SKK)
- Spanien (ESP)
- Südafrika (ZAR)
- Südkorea (KRW)
- Tschechien (CZK)
- Türkei (TRY)
- Ungarn (HUF)
- Vereinigtes Königreich (GBP)
- Vietnam (VND)

Zur Auswahl der Länder des angestrebten Datensatzes wurden andere, hier zitierte, empirische Arbeiten aus dem Devisenkontext gesichtet und eine Liste mit der Vereinigungsmenge der jeweils untersuchten Länder erstellt, da annahmegemäß bereits in einer anderen Studie verwendete Daten am ehesten verfügbar sein sollten. Dieser Prozess wurde nach den aufgeführten 42 Ländern beendet, um genügend zeitliche Ressourcen für die notwendige manuelle Bestimmung der Bloomberg-Ticker sowie die folgende Prüfung der Daten zu garantieren.

Nutzt man die IWF-Definitionen<sup>20</sup> der Begriffe Industriestaat und Entwicklungsland (IWF (2018)), so umfasst der angestrebte Datensatz 26 Industriestaaten und 16 Entwicklungsländer. Er beinhaltet Währungen bzw. Länder von allen<sup>21</sup> fünf verschiedenen Kontinenten.

Tabelle 1 fasst den generierten Datensatz als Übersicht zusammen. Als Quellen wurden Bloomberg, die OECD sowie der IWF einbezogen. Zusammenfassend ist die Datenlage beim Leistungsbilanzsaldo und der Produktionslücke am schlechtesten, da hier die geringsten Frequenzen und wenigsten Zeitreihen vorliegen. Für Marktdaten wie Spot- und Forward-Rates hingegen, aber auch für den Preisindex konnte ich bei fast allen Ländern Zeitreihen in der gewünschten monatlichen Frequenz erhalten.

Für detailliertere Informationen zum akquirierten Datensatz sei zuletzt auf die Übersichtstabellen in Anhang A und B verwiesen. Die Tabellen in Anhang A zeigen für die Zeitreihen jeder Datenquelle die Verfügbarkeit sowie ggf. den Namen des verwendeten Bloomberg-Tickers, das Start- sowie Enddatum, die abrufbare Datenfrequenz und die Sollanzahl der Beobachtungen seit 1974. Letztere wird zudem in weitere Komponenten aufgesplittet – die Anzahl der vorhandenen Beobachtungen, die Anzahl der fehlenden Beobachtungen und die Anzahl der fehlenden Beobachtungen nach Start der Reihe, d.h. nachdem die Zeitreihe mit ihrem ersten Wert begonnen hat. Ist der zuletzt genannte Wert hoch, kann dies entweder durch eine geringere als die angestrebte monatliche Datenfrequenz oder Datenprobleme wie unstete Zeitreihen rationalisiert werden. Der Wert stellt demnach einen Maßstab für die Vollständigkeit der Zeitreihe dar und soll als Indikator für eine etwaige händische Überprüfung der Reihe auf Tauglichkeit zur weiteren analytischen Verwendung dienen. Um Fälle, bei denen sich ein hoher Wert aufgrund der leider nicht beeinflussbaren Frequenz ergibt, von den anderen Fällen, die evtl. Datenprobleme aufweisen und händisch überprüft werden sollen, abzugrenzen, findet sich in der letzten Zeile der Tabelle zum Abgleich die Anzahl der erwarteten fehlenden Beobachtungen, gegeben dem Startdatum und der Frequenz der Zeitreihe. Übersteigt der tatsächliche Wert die Erwartung wesentlich, sollte eine händische Prüfung der Reihe erfolgen. Die Tabelle aus Anhang B liefert pro Datentyp und Währung den auf zwei Kommastellen gerundeten Pearson-Korrelationskoeffizienten, falls zwei vergleichbare Zeitreihen zur Verfügung stehen. Zusammenfassend kann eine vergleichsweise hohe Korrelation konstatiert werden.

#### 4.2. Vorgehen bei spezifischen Fragestellungen zur Datenverwendung

In diesem Abschnitt sollen die folgenden vier Punkte diskutiert werden:

- Fehlende Datenpunkte innerhalb der Zeitreihe
- Daten zu Ländern der Eurozone
- Zeitverzögerte Veröffentlichung von Daten
- Datenrevision

Die Gründe für fehlende Werte innerhalb der Zeitreihen können zweierlei sein: Zum einen mag die Frequenz der vorliegenden Daten geringer sein als die gewünschte Frequenz, zum anderen mag die Zeitreihe tatsächlich vom Datenanbieter unregelmäßig bestückt worden sein; natürlich ist auch beides denkbar. Unabhängig vom jeweiligen Grund sollte ein für alle Daten konsistentes Vorgehen zum Umgang mit solchen Fällen festgelegt werden. Hierbei orientiere ich mich an Barroso und Santa-Clara (2015), welche im Zuge von quartalsweisen CPI-Daten den vorherigen Wert in die nächste Periode fortschreiben.

<sup>20</sup>Im World Economic Outlook des IWF von April 2018 wird erklärt, dass die Klassifizierung nicht auf strikten Kriterien beruht, sondern mit Blick auf eine möglichst sinnvolle Einteilung der Daten zur Erleichterung der weiteren Analyse gewählt wird (IWF (2018)).

<sup>21</sup>Die Aussage basiert auf der Unterstellung eines Modells mit fünf Kontinenten, namentlich Amerika, Europa, Asien, Afrika und Australien.



**Tabelle 1: Generierter Datensatz**

Die Tabelle stellt den generierten Datensatz dar, aufgeschlüsselt nach den unterschiedlichen benötigten Datentypen. Sämtliche Daten wurden über ein Bloomberg-Terminal bezogen, sieht man vom risikolosen Zinssatz ab, welcher von der Internetseite des Ökonomen Kenneth French stammt, zu finden unter folgendem Link: [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). Die dritte Spalte gibt an, ob eine tägliche (tägl.), monatliche (monatl.), vierteljährliche (viertelj.) oder jährliche (jährl.) Frequenz vorliegt. Mehrere Angaben in dieser Spalte bedeuten, dass für die genannten Frequenzen jeweils mindestens zwei Zeitreihen vorhanden sind. Die letzte Spalte gibt pro Quelle Auskunft, für wie viele der 42 Länder des angestrebten Datensatzes Zeitreihen vorhanden sind, unabhängig von deren zeitlicher Länge.

| Datentyp                                                        | Datenquelle    | Vorhandene Frequenz(en) | Vorhandene Zeitreihen |
|-----------------------------------------------------------------|----------------|-------------------------|-----------------------|
| Mid-Spot Rates                                                  | Bloomberg      | Monatl./Tägl.           | 42 / 42               |
| Mid-1-Monat-Forward Rates                                       | Bloomberg      | Monatl.                 | 41 / 42               |
| Preisindex der Länder                                           | IWF (CPI)      | Monatl./ Viertelj.      | 42 / 42               |
|                                                                 | OECD (CPI)     | Monatl./ Viertelj.      | 36 / 42               |
| Leistungsbilanzsaldo in Prozent des BIPs                        | IWF            | Jährl.                  | 42 / 42               |
|                                                                 | OECD           | Viertelj.               | 34 / 42               |
| Produktionslücke in Prozent des BIPs                            | IWF            | Jährl.                  | 22 / 42               |
|                                                                 | OECD           | Jährl.                  | 30 / 42               |
| Risikoloser US-Zinssatz<br>(T-Bills) für einmonatige Haltedauer | Kenneth French | Monatl.                 | -                     |

Des Weiteren umfasst der angestrebte Datensatz zwölf Länder, deren eigene Währungen im Laufe der Zeit vom Euro ersetzt wurden. In Belgien, Deutschland, Finnland, Frankreich, Irland, Italien, Portugal, Österreich und Spanien wurde die Gemeinschaftswährung zu Beginn des Jahres 1999 eingeführt, während Griechenland zum ersten Januar 2001 und die Slowakei zum ersten Januar 2009 nachzogen. Das in der Methodologie beschriebene Universum von  $N_t$  Währungen, welches zum Formen eines Portfolios zu jedem Zeitpunkt  $t$  zur Verfügung steht, soll diese Termine berücksichtigen. Somit müssen zum Beginn des Jahres 1999 die Währungen von zehn Ländern durch den Euro ersetzt werden, während die griechische Drachme und die slowakische Krone zu ihren jeweiligen Einführungsterminen das Währungsuniversum verlassen.

Im Gegensatz zu Kapitalmarktdaten können makroökonomische Daten wie CPI, Leistungsbilanzsaldo oder die Produktionslücke typischerweise weder direkt zum Stichtag noch final, d.h. ohne spätere Anpassungen durch den Herausgeber aufgrund verbesserter Datenlage, festgestellt werden. Barroso und Santa-Clara (2015) treffen daher im Zuge ihrer Portfoliooptimierung die Annahme, dass der Leistungsbilanzsaldo des vorherigen Jahres im April des aktuellen Jahres bekannt wird.<sup>22</sup> Dies entspricht einer Verzögerung von vier Monaten, welche auch in der vorliegenden Arbeit sowohl für Leistungsbilanzsaldo als auch Produktionslücke unterstellt wird. Bei den CPI-Daten wird ebenfalls die Vorgehensweise von Barroso und Santa-Clara (2015) repliziert, welche eine Verzögerung von zwei Monaten unterstellen, um sicherzugehen, dass die Informationen zum jeweiligen Zeitpunkt der Betrachtung bekannt waren.

Zudem sollte ein möglicher Vorbehalt der weiteren Datenanalyse mit Blick auf die Revision von Daten erwähnt werden. Für einen möglichst realistischen Test des Optimierungs-

verfahrens mit den selektierten Charakteristika wäre es wünschenswert, dem hypothetischen Investor lediglich solche Informationen zur Hand zu geben, welche ein tatsächlicher Investor seinerzeit zur Entscheidungsfindung hätte heranziehen können. Dies steht im Widerspruch zu den vorliegenden Daten von IWF und OECD, welche nach Zugewinn neuer Erkenntnisse nachträglich angepasst wurden. Einige Paper, aufgelistet durch Croushore und Stark (2003), beschäftigen sich mit dem Einfluss von Revisionen auf die Resultate von Datenanalysen und weisen darauf hin, dass dem Thema durchaus Relevanz beigemessen werden sollte. Da allerdings schlichtweg kein dahingehender Datensatz vorliegt, wird im Folgenden von dieser Thematik abstrahiert.

#### 4.3. Selektion und Vorstellung des finalen Datensatzes

Die Bestimmung des finalen Datensatzes erfolgte, indem die Voraussetzung formuliert wurden, dass pro Land und Datentyp mindestens 25 Jahre an Daten zur Verfügung stehen müssen, unabhängig davon in welcher Frequenz diese Daten vorliegen oder von welcher Quelle sie bezogen wurden. Dies führte zu den folgenden verbliebenen 19 Ländern<sup>23</sup>

- Australien (AUD)
- Belgien (BEF)
- Dänemark (DKK)
- Deutschland (DEM)
- Eurozone (EUR)
- Finnland (FIM)
- Frankreich (FRF)
- Griechenland (GRD)
- Italien (ITL)

<sup>23</sup>Portugal wies zwar ebenfalls für alle Datentypen 25 Jahre an Daten auf, allerdings reichten die Spot-Zeitreihen nicht weit genug in die Vergangenheit, um die langfristige Trendumkehr zu berechnen, wie sie in Abschnitt 3.3.3 definiert wurde.

<sup>22</sup>Diese Aussage bezieht sich auf ihren Datensatz mit jährlicher Frequenz.

- Japan (JPY)
- Kanada (CAD)
- Neuseeland (NZD)
- Niederlande (NLG)
- Norwegen (NOK)
- Österreich (ATS)
- Schweden (SEK)
- Schweiz (CHF)
- Spanien (ESP)
- Vereinigtes Königreich (GBP)

Danach wurde der maximal vorhandene Zeitraum für diese Länder unter Bezugnahme aller Quellen überprüft und das Startdatum des Datensatzes zum Jahresende 1989 identifiziert. Das Ende des Samples wurde auf Ende 2017 datiert, da die Zeitreihen zu Leistungsbilanzsaldo und Produktionslücke nur bis dahin vorliegen.

Als nächstes folgte pro Datentyp und Währung die Prüfung, welche Datenquelle die meisten vorhandenen monatlichen Beobachtungen liefert. Die Anzahl der Beobachtungen einer Zeitreihe hängt dabei von ihrer Zeitabdeckung und Datenfrequenz ab. Wenn eine identische Anzahl an Beobachtungen vorlag, wurde arbiträr die IWF-Zeitreihe bevorzugt. Die selektierten und in Tabelle C.20 des Anhangs C dargestellten Zeitreihen wurden bei der Zusammensetzung des finalen Datensatzes bevorzugt, vorausgesetzt sie standen für die besagten 19 Länder über die definierte Zeitspanne seit 1989 zur Verfügung. Ansonsten habe ich die Zeitreihen der alternativen Quelle verwendet.

Nach der zuvor definierten Adjustierung aufgrund der zeitverzögerten Veröffentlichung von Daten beginnt die finale Betrachtungsperiode Ende April 1990 und schließt mit dem 29.12.2017 ab. Dabei liegen zeitgleich maximal 18 Währungen vor, bevor Anfang 1999 acht Währungen durch den Euro ersetzt werden und bis schließlich Anfang 2001 auch Griechenland die Gemeinschaftswährung einführt. Im Vergleich zu Barroso und Santa-Clara (2015), deren Datensatz zeitgleich maximal 21 und minimal 13 Währungen bzw. Länder aufweist, ist das Währungsuniversum des selektierten Datensatzes mit maximal 18 und minimal zehn Währungen geringer. Zudem deckt er mit 27 Jahren gegenüber Barroso und Santa-Clara (2015) und ihrer Gesamtbetrachtungsperiode von 02/1976 bis 12/2011 eine geringere Zeitspanne ab, beinhaltet aber aktuellere Daten.

## 5. Ergebnisse

### 5.1. Aufteilung des Datensatzes und Bildung von Kombinationsstrategien

Für die weitere Analyse wird der vorgestellte Gesamtdatensatz wie in Barroso und Santa-Clara (2015) zeitlich zweigeteilt. Zunächst erfolgt der In-Sample-Test, welcher die Charakteristika identifizieren soll, deren Relevanz der hypothetische Investor zum Endzeitpunkt der In-Sample-Periode bereits hätte erkennen können. Mit diesen Informationen wird

anschließend der Out-of-Sample-Test durchgeführt, bei welchem der Investor die Portfoliogewichte zu jedem Zeitpunkt lediglich auf Basis bis dahin vorhandener Informationen bestimmt. Demnach werden zu jedem Zeitpunkt die Koeffizienten unter Berücksichtigung aller bis dahin verfügbarer Datenpunkte geschätzt, um anschließend mittels dieser Koeffizienten und den aktuellen Charakteristika-Werten die optimalen Portfoliogewichte für die kommende Periode zu berechnen.

Die Wahl der jeweiligen Länge beider Testperioden wurde von Barroso und Santa-Clara (2015) inspiriert, indem ein ähnliches Verhältnis von Perioden- zu Gesamtdatensatzlänge Verwendung findet. Somit resultiert die Aufteilung in die In-Sample-Periode von April 1990 bis Oktober 2005 und die Out-of-Sample-Periode von November 2005 bis Dezember 2017.

Damit endet die In-Sample-Periode vor dem Beginn der globalen Finanzkrise, welche das darauffolgende, maßgeblich durch die amerikanische und europäische Notenbank getriebene Niedrigzinsniveau einläutete. Die Analyse wird zeigen, wie die anhand der historischen Daten optimierten Währungsportfolios des beschriebenen Währungsuniversums in der besagten Krisenzeit und dem neuen Wirtschaftsumfeld performt haben.

In der weiteren Analyse wird zwischen optimierten Einzelstrategien und Kombinationsstrategien unterschieden. Eine Einzelstrategie bezeichnet dabei das Resultat einer Optimierung, welche nur mit einer einzelnen Charakteristik als Inputfaktor arbeitet, während Kombinationsstrategien auf mehr als einem Signal beruhen. Um ein besseres Verständnis dafür zu erlangen, wie das Hinzufügen von Charakteristika das Ergebnis des Verfahrens verändert bzw. ob diese einen Mehrwert liefern, wurde der Optimierung sukzessiv immer eine weitere Charakteristik als Inputfaktor hinzugefügt, bis die finale Strategie alle vorhandenen Charakteristika miteinbezieht. Die Reihenfolge orientiert sich dabei grob am Grad der empirischen Dokumentation und theoretischen Fundierung der jeweiligen Charakteristik in der Literatur, wie in den Abschnitten 1.1 und 3.2 vorgestellt. Die lange und ausführliche Abhandlung von Carry in der Literatur verschafft diesem Signal nicht nur die Wahl als Benchmark in der vorliegenden Arbeit, sondern führt auch zur ersten Strategiekombination zusammen mit Momentum, einem ebenfalls gut dokumentierten Faktor. Im Anschluss wird die langfristige Trendumkehr (realer Wechselkurs) ergänzt, ein ähnlich gut dokumentierter Faktor wie Momentum. Als nächstes folgt der Leistungsbilanzsaldo, welcher zwar bereits lange in der Literatur diskutiert wird, aber dessen Evidenz als umstrittener bezeichnet werden kann. Zuletzt folgen mit der Produktionslücke und dem Volatilitätsfaktor jene zwei Charakteristika, die erst in den letzten Jahren an verstärkter Aufmerksamkeit gewonnen haben. In Abhängigkeit von den jeweiligen Ergebnissen dieser Einzel- und sukzessiven Kombinationsstrategien werden im weiteren Verlauf der Arbeit an gegebener Stelle zusätzliche Kombinationsstrategien eingeführt und untersucht.

## 5.2. In-Sample-Ergebnisse

### 5.2.1. Performance

Tabelle 2 präsentiert die Performance der optimierten Portfolios sowie jene der definierten Benchmarks über die Zeitperiode von April 1990 bis Oktober 2005. Die höchste monatliche Sharpe Ratio unter den Einzelstrategien kann das auf Basis der Carry-Charakteristik optimierte Portfolio mit 0,35 erzielen. Relativiert wird diese aber zum einen von der diversifizierten Carry-Benchmark, welche ebenfalls eine Sharpe Ratio von 0,35 aufweist, sowie von einem bemerkenswerten Crash-Risiko. Letzteres manifestiert sich in einer Schiefe von -0,60 sowie einer Exzess-Kurtosis von 1,11. Zudem weist die Carry-Einzelstrategie unter den Einzelstrategien mit Abstand den höchsten Leverage und Umschlag auf.

Mit der zweithöchsten Sharpe Ratio von 0,23 kann die auf der langfristigen Trendumkehr fundierte Einzelstrategie aufwarten. Zwar haftet der Wahrscheinlichkeitsverteilung dieses Portfolios mit einem Wert von 1,48 die höchste Exzess-Kurtosis aller betrachteten Einzelstrategien an, interessanterweise geht diese aber mit einer positiven Schiefe von 0,16 einher. Reiht man die Einzelstrategien weiter nach ihrer Sharpe Ratio, steht mit einer risikoadjustierten Performance von 0,09 die Momentum-Strategie an dritter Stelle. Bemerkenswert ist, dass dieses Portfolio die höchste positive Schiefe (0,22) verzeichnet.

Die von der Produktionslücke abhängige Strategie folgt als nächstes mit einer Sharpe Ratio von 0,07, bevor sich die Volatilitätsstrategie mit einem Wert von 0,03 eingliedert. Als Schlusslicht reiht sich die Leistungsbilanz-Charakteristik mit einer Sharpe Ratio von 0,02 ein. Wenig überrascht, dass die Produktionslücke- und Leistungsbilanzsaldo-Einzelstrategien den geringsten Umschlag unter allen untersuchten Strategien (abgesehen von der naiven Diversifikation) aufweisen, da die nötigen Daten nur mit quartalsweiser oder jährlicher Frequenz vorliegen, daher gemäß dem in Abschnitt 4.2 definierten Vorgehen öfters in die nächste Periode fortgeschrieben werden und als Signal somit weniger Variabilität bieten.

Wirft man einen Blick auf die erste der Kombinationsstrategien, welche Carry und Momentum vereint, stellt man zunächst fest, dass sich die Sharpe Ratio im Vergleich zur Carry-Einzelstrategie nur geringfügig um 0,01 auf 0,36 verbessert hat. Allerdings konstatiert ein Blick auf Schiefe (-0,22) und Kurtosis (0,56) des Kombinationsportfolios im Vergleich zur Carry-Einzelstrategie mit einer Schiefe (-0,60) und Kurtosis (1,11), dass ein Rückgang des erwähnten Crash-Risikos zu verzeichnen ist. Dies schlägt sich im Vergleich zur Carry-Einzelstrategie auch in einer leichten Erhöhung des monatlichen Sicherheitsäquivalents um 0,1% nieder. Letzteres definiert sich als die risikolose konstante Rendite, welche zum gleichen Nutzenniveau wie die Renditen der jeweiligen Strategie führt. Da die Kennzahl somit Abhängigkeit zur ausgewählten Nutzenfunktion aufweist, bezieht sie im vorliegenden Fall der CRRA-Nutzenfunktion auch die höheren Momente der Wahrscheinlichkeitsverteilung ein.

Ergänzt man als nächstes die langfristige Trendumkehr als Inputfaktor, verändert sich das Bild stärker. So steigt nicht nur die monatliche Sharpe Ratio um 0,07 relativ zur vorher beschriebenen Strategie auf 0,43, sondern weist die Strategie nun eine positive Schiefe (0,26) sowie eine mehr als verdreifachte Exzess-Kurtosis (1,91) auf. Zusammen resultiert dies in einem Anstieg des Sicherheitsäquivalents von ca. 0,82%.

Ein Hinzufügen der Leistungsbilanz-Charakteristik erhöht die Sharpe Ratio weiter um 0,03 auf 0,46, ergänzt positive Schiefe und führt zu einer Verbesserung des Sicherheitsäquivalents um zusätzliche knapp 0,4% im Vergleich zur Strategie des vorherigen Absatzes. Die Produktionslücke als weiteres Renditesignal bereichert das optimierte Portfolio hingegen kaum – die Sharpe Ratio ändert sich nicht und das Sicherheitsäquivalent steigt lediglich um 0,01%.

Als letztes findet der Volatilitätsfaktor Einzug in die Optimierung. Die Sharpe Ratio steigt um 0,03 auf 0,49, sowohl herbeigeführt durch einen Anstieg in der durchschnittlichen Rendite von 5,7% auf 6,02% als auch einer leichten Senkung der Standardabweichung von 11,63% auf 11,54%. Das Sicherheitsäquivalent erhöht sich um weitere 0,22% auf 3,32%. Bemerkenswert ist hier die Verringerung der Schiefe von 0,36 auf 0,07, konsistent mit der Tatsache, dass von den Einzelstrategien nur Carry eine noch negativere Schiefe als der Volatilitätsfaktor besitzt.

Der Leverage nimmt im Zuge des sukzessiven Hinzufügens aller sechs Charakteristika stetig zu und erreicht bei der finalen Kombinationsstrategie einen Wert von 13,72, den höchsten unter allen untersuchten Strategien. Beim Umschlag hingegen ist das Bild nicht gleichsam eindeutig. So weist die Strategie auf Basis von Carry, Momentum und der langfristigen Trendumkehr mit 9,36 den geringsten Umschlag unter den sukzessiven Kombinationsstrategien auf.

Da sich aufgrund der bis hierher vollzogenen Analyse eine Relevanz von Carry, der langfristigen Trendumkehr und Momentum herauskristallisiert hat, wurden zur besseren Entscheidungsfindung bzgl. einer potentiellen vierten Charakteristik für den Out-of-Sample-Test zudem zwei zusätzliche Strategiekombinationen untersucht. Die erste bezieht zusätzlich zu den drei genannten Charakteristika die Produktionslücke, die zweite den Volatilitätsfaktor mit ein. In Bezug auf die Sharpe Ratio können jedoch beide (0,43 und 0,44) nicht mit der Strategie mithalten, welche zusätzlich zu den drei Faktoren den Leistungsbilanzsaldo betrachtet (0,46). Gleiches gilt auch bei der Performance-Evaluation via Sicherheitsäquivalent. Allerdings sei wiederum die geringere Schiefe vermerkt, zu welcher der Einbezug des Volatilitätsfaktors (0,2) im Vergleich zum Leistungsbilanzsaldo (0,32) und der Produktionslücke (0,29) führt.

Wirft man abschließend einen vergleichenden Blick auf die Benchmarks naive Diversifikation und US-Aktienmarkt, können zwei Beobachtungen formuliert werden. Zum einen, dass innerhalb des In-Sample-Datensatzes die Verwendung eines Inputsignals bei der Konstruktion der Devisenportfolios i.d.R. einen positiven Effekt auf die Sharpe Ratio und das Sicherheitsäquivalent hatte. Demnach weisen alle optimierten Portfolios, außer dem auf Basis der Leistungsbilanz-

**Tabelle 2:** In-Sample-Performance

Die Tabelle stellt die Performance-Kennzahlen der optimierten Strategien sowie der Benchmarks über den Zeitraum der In-Sample-Periode von April 1990 bis Oktober 2005 dar. Die Optimierung betrachtet eine CRRA-Nutzenfunktion mit  $\gamma = 4$ . Aufgeführt werden in den Spalten die maximale ( $\max r$ ), minimale ( $\min r$ ) sowie durchschnittliche ( $\bar{r}$ ) monatliche Rendite, die geschätzte monatliche Standardabweichung ( $\sigma(r)$ ) (alle bisherigen Angaben in Prozent), Exzess-Kurtosis und Schiefe der Wahrscheinlichkeitsverteilung, die monatliche Sharpe Ratio (SR) und das monatliche Sicherheitsäquivalent in Prozent (SE). Die letzten beiden Spalten geben Auskunft über den durchschnittlichen Leverage ( $\bar{L}$ ), angegeben als der Betrag, welcher pro in die risikolose Geldanlage investiertem Dollar in das riskante Währungsportfolio angelegt ist, und Umschlag ( $\bar{T}$ ), der sich als die Summe aus den absoluten Änderungen aller Währungsgewichte des Forward-Portfolios zwischen zwei Perioden definiert. In den ersten sechs Zeilen werden die Kennzahlen für die Einzelstrategien abgebildet. Eine Einzelstrategie bezeichnet dabei das Resultat einer Optimierung, welche nur mit einer einzelnen Charakteristika als Inputfaktor arbeitet. Namentlich sind das folgende Strategien: Carry (fd), Momentum (mom), langfristige Trendumkehr (q), Leistungsbilanzsaldo (ca), die Produktionslücke (opg) sowie der Volatilitätsfaktor (vol). In den nächsten fünf Zeilen folgt die Betrachtung von Strategiekombinationen. Hierfür wurde der Optimierung in einer arbiträren Reihenfolge sukzessiv immer eine weitere Charakteristika als Inputfaktor hinzugefügt, bis die finale Strategie alle vorhandenen Charakteristika miteinbezieht. Wegen der arbiträren Reihenfolge finden sich in den folgenden zwei Zeilen zwei zusätzliche Strategiekombinationen, welche die drei aus der vorangegangenen Analyse relevantesten Charakteristika Carry, Momentum und langfristige Trendumkehr mit jeweils der Produktionslücke und dem Volatilitätsfaktor paaren. In den letzten drei Zeilen der Tabelle findet sich die Performance der drei Benchmarks. Für die Berechnung des Umschlages der US-Aktien Benchmark lagen keine Daten vor.

| Strategie     | $\max r$ | $\min r$ | $\bar{r}$ | $\sigma(r)$ | Kurtosis | Schiefe | SR   | SE   | $\bar{L}$ | $\bar{T}$ |
|---------------|----------|----------|-----------|-------------|----------|---------|------|------|-----------|-----------|
| 1. <i>fd</i>  | 25.38    | -27.54   | 3.02      | 7.60        | 1.11     | -0.60   | 0.35 | 1.77 | 9.15      | 8.96      |
| 2. <i>mom</i> | 7.29     | -5.19    | 0.53      | 2.24        | 0.57     | 0.22    | 0.09 | 0.43 | 2.02      | 1.56      |
| 3. <i>q</i>   | 24.06    | -21.64   | 1.69      | 5.82        | 1.48     | 0.16    | 0.23 | 1.02 | 5.93      | 1.41      |
| 4. <i>ca</i>  | 1.99     | -1.62    | 0.34      | 0.50        | 1.24     | -0.09   | 0.02 | 0.34 | 0.50      | 0.03      |
| 5. <i>opg</i> | 4.98     | -4.58    | 0.47      | 1.85        | 0.08     | -0.10   | 0.07 | 0.40 | 2.40      | 0.17      |
| 6. <i>vol</i> | 2.97     | -2.97    | 0.36      | 0.88        | 1.10     | -0.21   | 0.03 | 0.35 | 0.74      | 0.64      |
| 1 - 2         | 31.60    | -21.52   | 3.30      | 8.34        | 0.56     | -0.22   | 0.36 | 1.87 | 9.79      | 9.54      |
| 1 - 3         | 42.89    | -32.03   | 4.82      | 10.36       | 1.91     | 0.26    | 0.43 | 2.69 | 11.26     | 9.36      |
| 1 - 4         | 48.35    | -27.75   | 5.65      | 11.53       | 1.56     | 0.32    | 0.46 | 3.09 | 12.94     | 10.64     |
| 1 - 5         | 49.95    | -26.48   | 5.70      | 11.63       | 1.66     | 0.36    | 0.46 | 3.10 | 13.00     | 10.50     |
| 1 - 6         | 50.53    | -29.28   | 6.02      | 11.54       | 1.59     | 0.07    | 0.49 | 3.32 | 13.72     | 11.04     |
| 1 - 3 & 5     | 43.15    | -31.22   | 4.83      | 10.40       | 1.92     | 0.29    | 0.43 | 2.70 | 11.20     | 9.23      |
| 1 - 3 & 6     | 47.57    | -31.88   | 4.87      | 10.35       | 2.12     | 0.20    | 0.44 | 2.73 | 11.36     | 9.49      |
| Carry         | 37.64    | -41.29   | 3.73      | 9.63        | 2.95     | -0.63   | 0.35 | 1.48 | 11.00     | 10.47     |
| Naiv          | 6.08     | -7.32    | 0.41      | 2.28        | 0.23     | -0.13   | 0.03 | 0.30 | 1.00      | 0.01      |
| US-Aktien     | 11.22    | -15.65   | 0.96      | 4.27        | 0.74     | -0.58   | 0.15 | 0.58 | 0.00      | -         |

Charakteristik, eine gleichhohe oder höhere Sharpe Ratio als die naive Diversifikation auf. Bei Betrachtung des Sicherheitsäquivalents als Performance-Maß dominieren ausnahmslos alle optimierten Portfolios die besagte Benchmark, konsistent mit dem Fakt, dass die Kennzahl sowie das Optimierungsverfahren auf der gleichen CRRA-Nutzenfunktion beruhen und es sich um In-Sample-Ergebnisse handelt.

Zum anderen wird auf Ebene der Asset-Klassen deutlich, dass sich – zumindest in diesem ersten In-Sample-Test – die risikoadjustierten Renditen der Kombinationsstrategien in Form der monatlichen Sharpe Ratio durchaus mit denen einer diversifizierten US-Aktienstrategie (0,15) messen können bzw. diese wesentlich übersteigen. Allerdings muss sich der hypothetische Investor einer höheren Volatilität und Exzess-Kurtosis bewusst sein, die mit den Devisenportfolios im Vergleich zum Aktieninvestment tendenziell einhergehen. Die positive Schiefe hingegen sollte auf das Interesse des Aktieninvestors stoßen.

### 5.2.2. Koeffizienten und statistische Signifikanz

In Tabelle 3 sind zu jeder Strategie die geschätzten Koeffizienten sowie jeweils in Klammern darunter deren Signifikanz in Form von p-Werten abgebildet. Sämtliche Koeffizi-

enten – abgesehen von denen der Volatilität-Charakteristik – weisen ein positives Vorzeichen und somit einen Zusammenhang auf, welcher in Einklang mit der vorgestellten Literatur steht. Wie Brandt et al. (2009) anmerken, kann die Größenordnung der Koeffizienten untereinander verglichen werden, da die Charakteristika wie beschrieben im Querschnitt standardisiert wurden.

Bei Betrachtung der Koeffizienten der Einzelstrategien zeichnen sich auf dem 5%- sowie 1%-Signifikanzniveau von allen Charakteristika nur Carry und die langfristige Trendumkehrung als relevant ab. Dieses Resultat bleibt auch nach Signifikanzanalyse der Kombinationsstrategien unangetastet. Beide Charakteristika weisen stets die höchsten geschätzten Koeffizienten auf und haben somit den größten Einfluss auf die Gewichtung einer Währung im optimalen Währungsportfolio.

Auf dem 5%-Signifikanzniveau signalisiert bei den Einzelstrategien ebenfalls kein anderer Koeffizient Signifikanz – für die Kombinationsstrategien kann diese Aussage jedoch nicht getroffen werden. Demnach ist der Koeffizient für die Momentum-Charakteristik bei fast allen untersuchten Kombinationsstrategien signifikant. Interessanterweise zeigt sich der geschätzte Parameter für die Leistungsbilanz-



**Tabelle 3:** Geschätzte Koeffizienten und Signifikanz

Die Tabelle zeigt für jede optimierte Strategie die über den Zeitraum der In-Sample-Periode von April 1990 bis Oktober 2005 geschätzten Koeffizienten der Charakteristika sowie jeweils in Klammern darunter deren Signifikanz in Form von p-Werten. Namentlich handelt es sich bei den Charakteristika um folgende: Carry (*fd*), Momentum (*mom*), langfristige Trendumkehr (*q*), Leistungsbilanzsaldo (*ca*), die Produktionslücke (*opg*) sowie der Volatilitätsfaktor (*vol*). Die Optimierung betrachtet eine CRRA-Nutzenfunktion mit  $\gamma = 4$ . Die p-Werte wurden analog zu Barroso und Santa-Clara (2015) über ein Bootstrapping-Verfahren bestimmt. Im Zuge dieses Verfahrens wurden pro Optimierung 1.000 zufällige Datensätze durch Ziehen von Beobachtungen mit Zurücklegen aus dem originalen Datensatz generiert und anschließend für jeden dieser das Optimierungsverfahren erneut durchgeführt. Die Ergebnisse der Koeffizienten wurden auf zwei und jene der p-Werte auf drei Nachkommastellen gerundet. Die Symbole \*, \*\* und \*\*\* kennzeichnen Koeffizienten mit einem p-Wert  $\leq 0,10$ ,  $\leq 0,05$  bzw.  $\leq 0,01$ .

| Strategie     | <i>fd</i>           | <i>mom</i>        | <i>q</i>           | <i>ca</i>         | <i>opg</i>      | <i>vol</i>       |
|---------------|---------------------|-------------------|--------------------|-------------------|-----------------|------------------|
| 1. <i>fd</i>  | 12.06***<br>(0)     | -                 | -                  | -                 | -               | -                |
| 2. <i>mom</i> | -                   | 2.7<br>(0.139)    | -                  | -                 | -               | -                |
| 3. <i>q</i>   | -                   | -                 | 7.96***<br>(0)     | -                 | -               | -                |
| 4. <i>ca</i>  | -                   | -                 | -                  | 0.63<br>(0.405)   | -               | -                |
| 5. <i>opg</i> | -                   | -                 | -                  | -                 | 3.12<br>(0.141) | -                |
| 6. <i>vol</i> | -                   | -                 | -                  | -                 | -               | -1.02<br>(0.682) |
| 1 - 2         | 12.47***<br>(0)     | 2.73<br>(0.165)   | -                  | -                 | -               | -                |
| 1 - 3         | 11.54***<br>(0.001) | 6.58**<br>(0.024) | 8.12***<br>(0.002) | -                 | -               | -                |
| 1 - 4         | 13.48***<br>(0)     | 6.62**<br>(0.016) | 9.68***<br>(0.001) | 5.72**<br>(0.042) | -               | -                |
| 1 - 5         | 13.26***<br>(0)     | 6.61**<br>(0.026) | 10.07***<br>(0)    | 5.96**<br>(0.032) | 1.31<br>(0.445) | -                |
| 1 - 6         | 13.61***<br>(0)     | 5.79*<br>(0.055)  | 10.59***<br>(0)    | 7.59**<br>(0.018) | 2.78<br>(0.325) | -5.01<br>(0.95)  |
| 1 - 3 & 5     | 11.36***<br>(0)     | 6.5**<br>(0.031)  | 8.35***<br>(0.003) | -                 | 1.19<br>(0.422) | -                |
| 1 - 3 & 6     | 11.65***<br>(0)     | 7.29**<br>(0.031) | 7.98***<br>(0.002) | -                 | -               | -1.94<br>(0.738) |

Charakteristik ebenfalls für alle drei Kombinationsstrategien signifikant, in welcher sie als Inputfaktor dient.

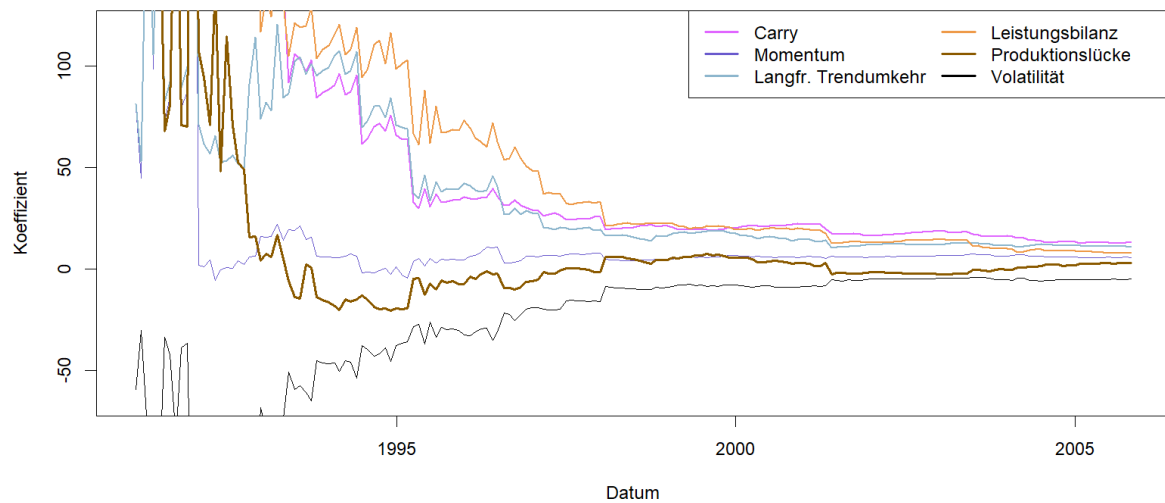
In Abbildung 1 habe ich graphisch dargestellt, wie sich die geschätzten Koeffizienten der untersuchten Charakteristika bei monatlicher Wiederholung des vorgestellten Verfahrens mit allen bis dahin vorhandenen Daten über den Zeitablauf der In-Sample-Periode verändern, wobei ich initial ein Jahr zum erstmaligen Schätzen der Koeffizienten verwendet habe. Es wird deutlich, dass das Optimierungsverfahren bei Verwendung des gegebenen Datensatzes einige Jahre braucht, um zu einigermaßen stabilen Koeffizienten zu gelangen. Gegen Anfang des Jahres 1998, etwa acht Jahre nach Start der Betrachtungsperiode, gehört die letzte vergleichsweise starke Änderung der Koeffizienten der Vergangenheit an, aber auch danach sind weiterhin Änderungen erkennbar. Diese sind aber i.d.R. weniger abrupt und gleichen in der Betrachtung eher einem Konvergenzprozess.

### 5.2.3. Einordnung und Interpretation der Ergebnisse

Die Ergebnisse des In-Sample-Tests sind in einem weiten Sinne gesprochen mit den Resultaten aus der wissenschaftlichen Literatur vereinbar, die in Abschnitt 1.1 und 3.2 vorgestellt wurden. Diese Aussage basiert auf der Beobachtung, dass innerhalb der In-Sample-Periode von April 1990 bis Oktober 2005 jene Charakteristika ökonomisch sowie statistisch besser performen, die bereits vergleichsweise lange und in einer gewissen Breite mit wissenschaftlicher Aufmerksamkeit bedacht wurden. Hier sind etwa Carry, Momentum sowie die auf dem realen Wechselkurs basierende langfristige Trendumkehr zu nennen, welche bereits ausgiebig in der FX-Literatur diskutiert wurden und innerhalb des In-Sample-Tests sowohl die ökonomische als auch statistische Relevanz dominieren.

Bei Betrachtung der übrigen drei Charakteristika, gilt die im letzten Absatz formulierte Beobachtung weiterhin, wenn auch weniger eindeutig. Für den Koeffizienten der Leistungsbilanz-Charakteristik, deren Einfluss auf Währungs-





**Abbildung 1:** Veränderung der Koeffizienten im Zeitablauf der In-Sample-Periode. Die Abbildung stellt die zu den Zeitpunkten des In-Sample-Tests (April 1990 bis Oktober 2005) auf Basis aller bis dahin vorhandener Daten geschätzten Koeffizienten der fünften und letzten Kombinationsstrategie dar, welche alle definierten Charakteristika berücksichtigt. Für den ersten Datenpunkt der Abbildung wurden zwölf Monate an Daten zum erstmaligen Schätzen der Koeffizienten verwendet.

renditen ebenfalls bereits vergleichsweise lange von wissenschaftlichen Arbeiten untersucht wird<sup>24</sup>, kann in drei Kombinationsstrategien Signifikanz auf dem 5%-Niveau nachgewiesen werden, während für Produktionslücke und Volatilität, die als relevante Faktoren in der Forschung erst später zu Aufmerksamkeit kamen, in keinem der Strategiesetups Signifikant festzustellen ist. Es muss an dieser Stelle jedoch auch bemerkt werden, dass die ökonomische Performance – sowohl gemessen an Sharpe Ratio als auch Sicherheitsäquivalent – der Leistungsbilanz von diesen drei Charakteristika am geringsten ist.

Beim Verlassen dieser recht allgemeinen Interpretationsebene fällt als weitere Parallele zur vorgestellten wissenschaftlichen Literatur auf, dass sowohl Carry-Einzelstrategie als auch Carry-Benchmark ein erhebliches Crash-Risiko aufweisen, wie es bspw. schon Brunnermeier et al. (2008) feststellten. Ebenso kann im Zuge des In-Sample-Tests die Aussage von Jordà und Taylor (2012) bestätigt werden, welche die Kombination einer Carry-Strategie mit einem Signal basierend auf dem realen Wechselkurs als Möglichkeit beschreiben, Diversifikationsvorteile zu erzielen und das besagte Crash-Risiko des Carry-Trades zu senken. Hier lassen sich auch Analogien zu der Arbeit von Giordani und Halling (2019) ableiten, welche im Aktienkontext einen positiven Zusammenhang zwischen Bewertungsniveaus und negativen Extremereignissen feststellen. Auch Momentum verspricht gemäß den präsentierten In-Sample-Ergebnissen in diesem Sinne Minderung des Crash-Risikos.

<sup>24</sup>Anzumerken ist, dass der Einfluss des Leistungsbilanzsaldos zwar – wie in Abschnitt 3.2.3 dargestellt – seit geraumer Zeit untersucht wird, die empirische Evidenz für diesen Einfluss allerdings umstrittener ist als bei bspw. Carry, Momentum und dem realen Wechselkurs.

Wie in Abschnitt 3.2 zu bemerken war, werden vielversprechende Diversifikationseigenschaften oft als ein Argument herangezogen, um die Vorteilhaftigkeit einer vorgeschlagenen Strategie zu unterstreichen. Die Ergebnisse einer durchgeführten Korrelationsanalyse, zu finden in Anhang D, bestätigen diese Diversifikationseigenschaften, da die verschiedenen Einzelstrategien innerhalb der In-Sample-Periode ein eher unkorreliertes bzw. negativ korreliertes Gesamtbild aufweisen. Betrachtet man die Korrelationen zwischen der Produktionslücke-Strategie und den übrigen verwendeten Einzelstrategien, stößt man für Carry (0,26) auf den höchsten Wert. Riddiough und Sarno (2017) dagegen attestieren in ihrer Analyse der Produktionslücke-Strategie für diese beiden Benchmarks eine vergleichsweise geringe Korrelation. Allerdings schränken unterschiedliche methodologische Definitionen des Faktors die Vergleichbarkeit zwischen der zitierten und der vorliegenden Arbeit ein.

Gegenüber der Carry-Benchmark bestätigt sich das eher unkorrelierte Gesamtbild, sieht man von den Carry- (0,96) und Produktionslücke-Einzelstrategien (0,22) ab. Für die  $\frac{1}{n}$ -Benchmark gilt das gleiche, wobei Leistungsbilanzsaldo (0,37) und Produktionslücke (0,14) die höchsten Korrelationen erreichen. Ebenfalls zeigt sich in Anbetracht der Korrelationen mit der US-Aktien-Benchmark, dass die implementierten internationalen Währungsstrategien einem US-Aktienbesitzer wesentliche Diversifikationsvorteile versprechen.

Wesentliches Ziel des In-Sample-Tests ist es festzustellen, welche Charakteristika ein hypothetischer Investor zum Ende der In-Sample-Periode mit dem untersuchten Datensatz als relevant hätte identifizieren können, um sie anschließend in einem Out-of-Sample-Experiment einzusetzen. Die prä-

sentierten Ergebnisse, bei denen Carry, die langfristige Trendumkehr und Momentum am besten abschnitten, sprechen für eine Berücksichtigung dieser Charakteristika im Out-of-Sample-Test. Dies stimmt, trotz Unterschieden im betrachteten Währungs- und Zeitumfang, mit dem In-Sample-Resultat von Barroso und Santa-Clara (2015) überein, welche den genannten Charakteristika ebenfalls Signifikanz zusprechen. Da in der vorliegenden Analyse bei zusätzlicher Betrachtung des Leistungsbilanzsaldos ein weiterer Sprung von etwa 0,03 in der monatlichen Sharpe Ratio und 0,4% im Sicherheitsäquivalent folgt und ich für den Koeffizienten dieser Charakteristik auf dem 5%-Niveau Signifikanz nachweisen kann, wird auch diese Charakteristik in den Out-of-Sample-Test inkludiert. Barroso und Santa-Clara (2015) hingegen können für den Leistungsbilanzsaldo keine Signifikanz konstatieren.

### 5.3. Out-of-Sample-Ergebnisse

#### 5.3.1. Performance

Tabelle 4 präsentiert die Performance der optimierten Portfolios sowie jene der definierten Benchmarks über die Out-of-Sample-Zeitperiode von November 2005 bis Dezember 2017. Schnell ist zu erkennen, dass sich die Resultate stark von der In-Sample-Periode unterscheiden:

Von den Einzelstrategien können nunmehr lediglich jene auf Basis von Carry und der langfristigen Trendumkehr eine leicht positive Sharpe Ratio (0,03 und 0,04) bieten, alle anderen weisen für die Kennzahl ein negatives Vorzeichen auf. Am meisten abgenommen hat relativ zur In-Sample-Periode die Performance der Carry-Einzelstrategie, welche einen absoluten Einbruch von 0,32 in ihrer Sharpe Ratio hinnehmen muss. Im Vergleich dazu muss die verwandte Carry-Benchmark zwar ebenfalls einen absoluten Einbruch der Sharpe Ratio von 0,28 verzeichnen, allerdings ist die verbliebene risikoadjustierte Performance mit 0,07 nach wie vor höher als die der Einzelstrategie.

Der maßgebliche Einflussfaktor auf diesen Einbruch in den Sharpe Ratios der Einzelstrategien liegt vermutlich im Rückgang der durchschnittlichen monatlichen Renditen relativ zur In-Sample-Periode, welchen alle Einzelstrategien in unterschiedlichem Ausmaße verbuchen. Am stärksten von dieser Entwicklung betroffen sind Carry mit einem relativen Renditerückgang von 2,75% und die langfristige Trendumkehr mit -1,41%, also jene Strategien, welche zuvor im In-Sample-Test am besten abschnitten; sie weisen unter den Einzelstrategien aber nach wie vor die höchsten durchschnittlichen Renditen auf. Im Durchschnitt über alle Einzelstrategien ist der Mittelwert der monatlichen Renditen um 0,97% zurückgegangen. Die Volatilität, um den zweiten Einflussfaktor auf die Sharpe Ratio zu benennen<sup>25</sup>, ist für fünf der sechs Einzelstrategien relativ zur In-Sample-Periode tatsächlich gesunken. Wie die Sharpe Ratios allerdings erkennen lassen, konnte diese Verringerung von im Durchschnitt 0,36% nicht den erwähnten Einbruch in den durchschnittlichen monatlichen Renditen kompensieren.

Blickt man als nächstes Performance-Maß auf das Sicherheitsäquivalent, ergibt sich ein ähnliches Bild. Allgemein ist das Niveau bei allen Einzelstrategien stark gesunken und für vier von sechs weist die Kennzahl ein negatives Vorzeichen auf. Zusätzlich zu den bereits skizzierten Einflussfaktoren der Sharpe Ratio spielt beim Sicherheitsäquivalent der extreme Anstieg der Exzess-Kurtosis eine Rolle, welcher bei allen Einzelstrategien verzeichnet werden kann und im Durchschnitt 5,98 beträgt. Dies gleicht einem Anstieg von ca. 642% im Vergleich zur durchschnittlichen Exzess-Kurtosis aller Einzelstrategien über die In-Sample-Periode (0,93) und wirkt auf einen risikoaversen Investor abschreckend, welcher Extremereignisse nicht schätzt.

Betrachtet man die Einzelstrategien im Vergleich zu den Benchmarks anhand der Sharpe Ratio, kann festgestellt werden, dass die Carry-Benchmark und der US-Aktienmarkt jede der Einzelstrategien mit Werten von 0,07 und 0,18 dominieren. Die naive Diversifikation bleibt zwar mit einer Sharpe Ratio von 0,03 hinter der langfristigen Trendumkehr mit 0,04 zurück, trotzdem kann sie bemerkenswerterweise mit der gleichen Sharpe Ratio wie die Carry-Einzelstrategie aufwarten und schlägt alle verbliebenen Einzelstrategien in dieser Disziplin.

Blickt man im nächsten Schritt auf die Performance der Kombinationsstrategien in Tabelle 4 und beginnt dabei mit der Mischung aus Carry und Momentum, kann eine leichte Verbesserung der monatlichen Sharpe Ratio von 0,01 im Vergleich zur Carry-Einzelstrategie auf 0,04 konstatiert werden. Dies geht mit einer Steigerung in der durchschnittlichen Rendite von 0,27% auf 0,39% und einem Anstieg in der Standardabweichung von 7,18% auf 7,87% einher. Bemerkenswert ist der Rückgang der Exzess-Kurtosis um 5,46 auf 1,68 und eine Verbesserung der Schiefe von -1,52 auf -0,84. Das Sicherheitsäquivalent ist schließlich um 0,1% weniger negativ.

Einen positiven Effekt hat der weitere Einbezug der langfristigen Trendumkehr auf die Performance des optimierten Portfolios. Die Sharpe Ratio steigt demnach im Vergleich zum Portfolio, welches nur Carry und Momentum berücksichtigt, um 0,07 auf 0,11. Die durchschnittliche Rendite steigt dabei von 0,39% auf 1,05% und erreicht in Folge den höchsten Wert unter allen untersuchten Strategien, jedoch nimmt auch das Risiko gemessen an der Volatilität um 1,02% zu. Die Exzess-Kurtosis wächst um 0,62 auf 2,3 und weist auf eine Zunahme von Extremereignissen hin. Beachtenswert ist, dass die Schiefe um 1,47 angestiegen ist und nun einen positiven Wert in der Höhe von 0,63 innehat. Zusammen resultiert dies in einem Anstieg des Sicherheitsäquivalents um 0,59% – trotz allem ist letzteres mit einem Wert von -0,46% nach wie vor negativ.

Als nächstes wird der Leistungsbilanzsaldo in die Optimierung miteinbezogen. Diese Strategie stellt somit jene dar, welche der hypothetische Investor auf Basis der Überlegungen in Abschnitt 5.2.3 für sein Out-of-Sample-Ergebnis gewählt hat. Es stellt sich heraus, dass die vorherige Kombination aus Carry, Momentum und der langfristigen Trendumkehr ohne den Leistungsbilanzsaldo besser performt hat. Die Shar-

<sup>25</sup>Vom risikolosen Zins als Opportunitätskostensatz wird hier abstrahiert, da dieser exogen für alle Strategien identisch gegeben ist.

**Tabelle 4:** Out-of-Sample-Performance

Die Tabelle stellt die Performance-Kennzahlen der optimierten Strategien sowie der Benchmarks über den Zeitraum der Out-of-Sample-Periode von November 2005 bis Dezember 2017 dar. Die Optimierung betrachtet eine CRRA-Nutzenfunktion mit  $\gamma = 4$ . Aufgeführt werden in den Spalten die maximale ( $\max r$ ), minimale ( $\min r$ ) sowie durchschnittliche ( $\bar{r}$ ) monatliche Rendite, die geschätzte monatliche Standardabweichung ( $\sigma(r)$ ) (alle bisherigen Angaben in Prozent), Exzess-Kurtosis und Schiefe der Wahrscheinlichkeitsverteilung, die monatliche Sharpe Ratio (SR) und das monatliche Sicherheitsäquivalent in Prozent (SE). Die letzten beiden Spalten geben Auskunft über den durchschnittlichen Leverage ( $\bar{L}$ ), angegeben als der Betrag, welcher pro in die risikolose Geldanlage investiertem Dollar in das riskante Währungsportfolio angelegt ist, und Umschlag ( $\bar{T}$ ), der sich als die Summe aus den absoluten Änderungen aller Währungsgewichte des Forward-Portfolios zwischen zwei Perioden definiert. In den ersten sechs Zeilen werden die Kennzahlen für die Einzelstrategien abgebildet. Eine Einzelstrategie bezeichnet dabei das Resultat einer Optimierung, welche nur mit einer einzelnen Charakteristika als Inputfaktor arbeitet. Namentlich sind das folgende Strategien: Carry (fd), Momentum (mom), langfristige Trendumkehr (q), Leistungsbilanzsaldo (ca), die Produktionslücke (opg) sowie der Volatilitätsfaktor (vol). In den nächsten fünf Zeilen folgt die Betrachtung von Strategiekombinationen. Hierfür wurde der Optimierung in einer arbiträren Reihenfolge sukzessiv immer eine weitere Charakteristika als Inputfaktor hinzugefügt, bis die finale Strategie alle vorhandenen Charakteristika miteinbezieht. Wegen der arbiträren Reihenfolge finden sich in den folgenden zwei Zeilen zwei zusätzliche Strategiekombinationen, welche die drei aus der vorangegangenen Analyse relevantesten Charakteristika Carry, Momentum und langfristige Trendumkehr mit jeweils der Produktionslücke und dem Volatilitätsfaktor paaren. In den letzten drei Zeilen der Tabelle findet sich die Performance der drei Benchmarks. Für die Berechnung des Umschlages der US-Aktien Benchmark lagen keine Daten vor.

| Strategie     | $\max r$ | $\min r$ | $\bar{r}$ | $\sigma(r)$ | Kurtosis | Schiefe | SR    | SE    | $\bar{L}$ | $\bar{T}$ |
|---------------|----------|----------|-----------|-------------|----------|---------|-------|-------|-----------|-----------|
| 1. <i>fd</i>  | 17.91    | -40.28   | 0.27      | 7.18        | 7.14     | -1.52   | 0.03  | -1.15 | 5.26      | 5.85      |
| 2. <i>mom</i> | 11.05    | -6.58    | 0.04      | 2.30        | 3.08     | 0.44    | -0.02 | -0.06 | 1.73      | 1.38      |
| 3. <i>q</i>   | 28.80    | -17.35   | 0.28      | 5.02        | 6.87     | 1.07    | 0.04  | -0.20 | 4.02      | 0.92      |
| 4. <i>ca</i>  | 2.19     | -1.96    | 0.05      | 0.46        | 7.51     | -0.02   | -0.09 | 0.04  | 0.29      | 0.09      |
| 5. <i>opg</i> | 2.89     | -3.84    | -0.08     | 0.84        | 6.07     | -0.32   | -0.19 | -0.09 | 0.68      | 0.16      |
| 6. <i>vol</i> | 5.68     | -4.05    | 0.05      | 0.95        | 10.76    | 0.70    | -0.04 | 0.03  | 0.54      | 0.45      |
| 1 - 2         | 18.88    | -27.67   | 0.39      | 7.87        | 1.68     | -0.84   | 0.04  | -1.05 | 6.67      | 7.31      |
| 1 - 3         | 37.46    | -27.33   | 1.05      | 8.89        | 2.30     | 0.63    | 0.11  | -0.46 | 8.89      | 9.42      |
| 1 - 4         | 39.58    | -29.25   | 1.02      | 9.64        | 1.89     | 0.51    | 0.10  | -0.78 | 10.01     | 10.77     |
| 1 - 5         | 39.61    | -29.35   | 0.98      | 9.70        | 1.93     | 0.58    | 0.09  | -0.83 | 10.25     | 11.30     |
| 1 - 6         | 38.19    | -29.17   | 0.82      | 9.89        | 1.84     | 0.48    | 0.07  | -1.08 | 10.14     | 11.50     |
| 1 - 3 & 5     | 37.30    | -27.45   | 0.99      | 8.96        | 2.45     | 0.75    | 0.10  | -0.51 | 9.02      | 9.77      |
| 1 - 3 & 6     | 33.06    | -27.93   | 0.91      | 8.74        | 1.55     | 0.42    | 0.09  | -0.58 | 8.81      | 9.57      |
| Carry         | 34.44    | -48.97   | 0.82      | 9.80        | 4.59     | -0.83   | 0.07  | -1.92 | 7.69      | 8.52      |
| Naiv          | 7.5      | -8.86    | 0.16      | 2.58        | 1.01     | -0.16   | 0.03  | 0.03  | 1.00      | 0.01      |
| US-Aktien     | 11.35    | -17.15   | 0.83      | 4.17        | 2.02     | -0.79   | 0.18  | 0.46  | 0.00      | -         |

pe Ratio nimmt beim Hinzufügen des Saldos um 0,01 ab, was durch eine um 0,03% geringere mittlere Rendite und eine um 0,75% höhere Standardabweichung verursacht wird. Zwar sinkt des Weiteren die Exzess-Kurtosis um 0,41 auf 1,89, allerdings nimmt auch die Schiefe um 0,12 ab. Dies resultiert in einer Abnahme des Sicherheitsäquivalents um 0,32%. Diese Beobachtungen lassen sich auch für die folgenden zwei Kombinationsstrategien sowie die zwei untersuchten Zusatzstrategien machen – allerdings variiert jeweils die Stärke der Ausprägung. Wie im In-Sample-Test nimmt der Leverage im Zuge der sukzessiven Kombinationsstrategien in der Tendenz zu. Gleiches stelle ich nun auch für den Umschlag des Forward-Portfolios fest.

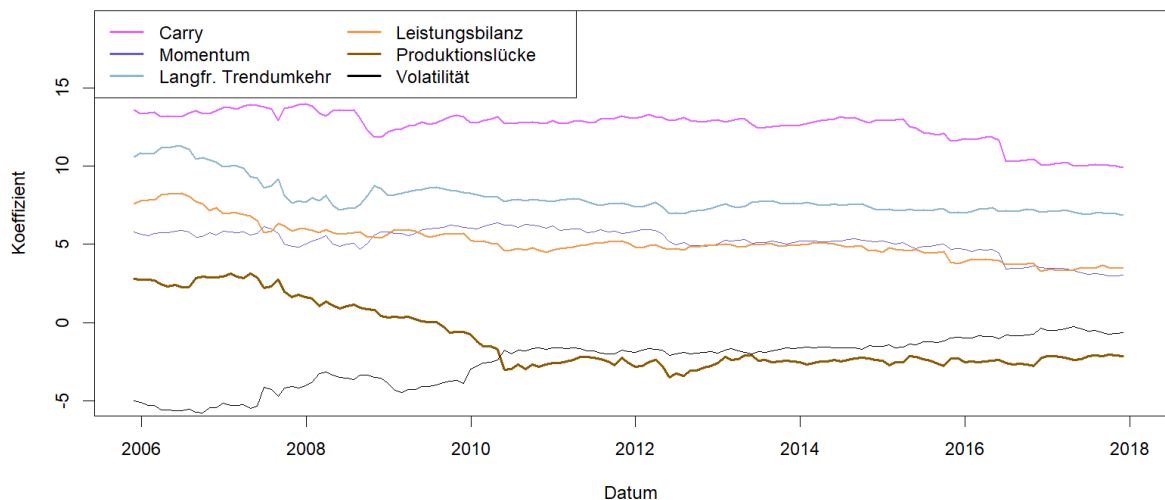
Zusammenfassend kann konstatiert werden, dass die Strategie auf Basis von Carry, Momentum und der langfristigen Trendumkehr von allen optimierten Strategien die höchste Sharpe Ratio bzw. das höchste Sicherheitsäquivalent erreicht. Im Vergleich dieser Strategie mit den Benchmarks kann sie zumindest die Carry-Benchmark und die naive Diversifikation sowohl im Hinblick der Sharpe Ratio als auch des Sicherheitsäquivalents schlagen. Allerdings bleibt auch sie hinter der US-Aktien-Benchmark zurück.

### 5.3.2. Veränderung der Koeffizienten im Zeitablauf

In Abbildung 2 wird graphisch dargestellt, wie sich die geschätzten Koeffizienten der untersuchten Charakteristika bei monatlicher Wiederholung des vorgestellten Verfahrens mit allen bis dahin vorhandenen Daten über den Zeitablauf der Out-of-Sample-Periode verändern.

Zunächst fällt auf, dass die stetige Erweiterung der verwendeten Daten vom Ende der In-Sample- bis zum Ende der Out-of-Sample-Periode i.d.R. zu einer Abnahme aller geschätzten Koeffizienten geführt hat – Ausnahme ist hierbei der Volatilitätsfaktor, welcher zunächst negativ ist und sich über die Betrachtungsperiode null annähert, folglich statistisch irrelevant wird. In Folge nimmt der Einfluss aller Charakteristika auf die Gewichtung der Währungen im optimalen Portfolio über die Zeit ab, da die definierten Faktoren über den Zeitraum weniger attraktiv auf den hypothetischen Investor mit der CRRA-Nutzenfunktion wirken.

Im Einzelnen sank mit Abstand am stärksten der Koeffizient der Produktionslücke-Charakteristika. Dieser fiel bereits in der zweiten Hälfte des Jahres 2009 unter null und konnte sich anschließend nicht mehr erholen. Als zweitgrößter Verlierer ist anschließend der Leistungsbilanzsaldo zu nennen. Der Koeffizient dieser Charakteristika nahm über die Out-



**Abbildung 2:** Veränderung der Koeffizienten im Zeitablauf der Out-of-Sample-Periode.

Die Abbildung stellt die zu jedem Zeitpunkt des Out-of-Sample-Tests (November 2005 bis Dezember 2017) auf Basis aller bis dahin vorhandener Daten geschätzten Koeffizienten der fünften und letzten Kombinationsstrategie dar, die alle definierten Charakteristika berücksichtigt.

of-Sample-Periode um etwa 55% ab, gefolgt von jenem der Momentum-Strategie mit ca. 48%. Am geringsten abgenommen haben die Koeffizienten der langfristigen Trendumkehr und von Carry, welche jeweils nur um 35% und 27% sanken.

Zu beobachten ist, dass ein Extremereignis wie die bereits erwähnte Endphase der Finanzkrise vor dem Jahresbeginn 2009 teils stark unterschiedlichen Einfluss auf die Koeffizienten der Charakteristika genommen hat. Während der geschätzte Koeffizient von Carry stark einbrach, verzeichnete Momentum einen geringeren Einbruch mit rascher Erholung. Parallel dazu blieb der Koeffizient der Leistungsbilanz-Charakteristik in etwa stabil, wohingegen der Koeffizient der langfristigen Trendumkehr sogar über den Zeitraum zunahm.

### 5.3.3. Einordnung und Interpretation der Ergebnisse

Die präsentierten Ergebnisse zeigen, wie In-Sample-Resultate von einem Out-of-Sample-Test relativiert werden. Keine optimierte Strategie kommt in Bezug auf Sharpe Ratio und Sicherheitsäquivalent auch nur ansatzweise in ähnliche Regionen wie zuvor im In-Sample-Test – tatsächlich weisen nun fast alle Sicherheitsäquivalente ein negatives Vorzeichen auf und keine der optimierten Strategien erreicht eine höhere monatliche Sharpe Ratio als 0,11. Ein Investment in die US-Aktien-Benchmark mit einer Sharpe Ratio von 0,18 wäre demnach attraktiver. Auch die naive Diversifikation wirkt plötzlich nicht mehr so naiv, kann sie doch in Sachen Sicherheitsäquivalent mit einem Wert von 0,03% alle optimierten Strategien mit Ausnahme der Leistungsbilanz-Einzelstrategie (0,04%) schlagen oder zumindest gleichziehen. Bei Betrachtung der Sharpe Ratio kann sie mit einem Wert von 0,03 zumindest gegen alle Einzelstrategien außer jener auf Basis der langfristigen Trendumkehr (0,04) bestehen.

Zur Verteidigung der Performance der optimierten Währungsstrategien sei erwähnt, dass die Carry-Benchmark ebenfalls einen massiven Performance-Einbruch erlitt. Es zeigt sich somit, dass die in Abschnitt 1.1 beschriebene Investitionsstrategie in der Out-of-Sample-Periode von November 2005 bis Dezember 2017 und gegeben dem zu diesem Zeitpunkt auf nur zehn Währungen von Industriestaaten geschrumpften Anlageuniversum geschwächt hat. In diesem Zusammenhang kann vergleichbar zum In-Sample-Test bestätigt werden, dass dem Carry-Trade im Allgemeinen ein starkes Crash-Risiko anhaftet, welches sich in der Schlussphase der Finanzkrise zu Beginn des Jahres 2009 massiv realisiert hat, wie schon durch [Jordà und Taylor \(2012\)](#) festgestellt.

Zudem findet sich wie im In-Sample-Test Beleg für die Aussage von [Jordà und Taylor \(2012\)](#), dass die Kombination einer Carry-Strategie mit einem Signal basierend auf dem realen Wechselkurs attraktive Diversifikationsvorteile bietet und zu einer Verminderung des Crash-Risikos führt. Demnach weist die Carry-Einzelstrategie und jene auf Basis der langfristigen Trendumkehr die negativste Korrelation (-0,55) zwischen allen untersuchten Strategien auf, zu finden als Teil der Korrelationsanalyse in Anhang D. Des Weiteren ergibt sich für die Renditeverteilung der Strategie auf Basis des realen Wechselkurses die höchste Schiefe (1,07) aller untersuchten Portfolios während der Out-of-Sample-Periode. Dementsprechend bringt im Zuge der sukzessiven Kombinationsstrategien das Hinzufügen dieser Charakteristik die höchste Steigerung in der Schiefe<sup>26</sup>. Zuletzt sei darauf verwiesen, dass

<sup>26</sup>Die Kombinationsstrategie auf Basis von Carry, Momentum und langfristiger Trendumkehr hat eine Schiefe von 0,63, während jene nur basierend auf Carry und Momentum eine negative Schiefe von -0,84 aufweist.



die Einzelstrategie der langfristigen Trendumkehr in der entscheidenden Phase der Finanzkrise, als der Carry-Trade kollabierte, mit Gewinnen aufwarten konnte.

Aber auch die Aussage von Menkhoff et al. (2012b), welche der FX-Momentum-Strategie Diversifikationsvorteile im Zusammenhang mit dem Carry-Trade zusprechen, erfährt durch die präsentierten In- sowie Out-of-Sample-Ergebnisse Unterstützung. So liegt die Korrelation zwischen der Momentum- und der Carry-Einzelstrategie im Zeitraum der Out-of-Sample-Periode bei -0,37, während die Korrelation zwischen Momentum und der Carry-Benchmark einen Wert von -0,33 aufweist. Im Vergleich zur In-Sample-Periode (0,08 und 0,04) haben sich die Korrelationen somit weiter verringert. Zudem vermag Momentum in der Kombination mit Carry ebenfalls das Crash-Risiko zu senken, wie eine Verbesserung der Schiefe (von -1,52 auf -0,84) und ein extremer Rückgang der Exzess-Kurtosis (von 7,14 auf 1,68) im Vergleich zur Carry-Einzelstrategie verdeutlichen.

Auch wenn es keine der untersuchten Strategien bei Betrachtung von Sharpe Ratio und Sicherheitsäquivalent mit jenen Portfolios aufnehmen kann, welche von Barroso und Santa-Clara (2015) über ihre Out-of-Sample-Periode von März 1996 bis Dezember 2012 unter Verwendung des gleichen Ansatzes für ein etwas größeres Währungsuniversum<sup>27</sup> getestet wurden, lässt sich trotzdem eine Gemeinsamkeit bei den Ergebnissen erkennen. So attestierten die Autoren unter ihren getesteten Charakteristika lediglich Carry, Momentum und der langfristigen Trendumkehr Signifikanz, während sie bspw. die Leistungsbilanzsaldo-Charakteristik ablehnten. In der vorliegenden Arbeit hat sich der hypothetische Investor nach Sichtung der In-Sample-Ergebnisse, welche dem Leistungsbilanzsaldo in Kombination mit Carry, Momentum und der langfristigen Trendumkehr Signifikanz zusprachen, zur Aufnahme dieser makroökonomischen Inputvariable in die optimale Strategie entschlossen. Wie sich allerdings im Zuge des anschließenden Out-of-Sample-Tests herausstellte, performte die Strategie auf Basis nur der drei genannten Charakteristika besser als bei zusätzlicher Betrachtung des Leistungsbilanzsaldos.

Interessant ist auch zuletzt der extreme Anstieg in der Korrelation mit der naiven Diversifikation, welchen Carry-Einzelstrategie (ca. 667%) sowie Carry-Benchmark (ca. 860%) im Vergleich zur In-Sample-Periode verzeichnen. Ein möglicher Interpretationsansatz steht mit der allgemeinen Niedrigzinsphase in den entwickelten Volkswirtschaften in Zusammenhang, welche sich an die globale Finanzkrise angliederte. Da als Resultat der Zinsunterschied zwischen den entwickelten Volkswirtschaften geschrumpft ist und somit im Querschnitt der Währungen weniger variiert, variieren dementsprechend die auf Basis dieses Signals definierten Portfoliogewichte weniger und weisen tendenziell mehr Ähnlichkeit zu einer gleichgewichteten Strategie auf. Anders gesprochen verliert das Carry-Signal durch die Niedrigzin-

sphase an Aussagekraft bei einem relativen Vergleich unter Industrienationen.

#### 5.4. Kritik am angewendeten Ansatz

Die bisher vollzogene Analyse gibt aus mehreren Gründen Anlass zu Kritik. Zum Ersten erfolgte etwa die Selektion des untersuchten Länder- bzw. Währungsumfanges aus der ursprünglichen Anzahl von 42 Ländern auf Basis der arbiträren Voraussetzung, dass pro Land und Datentyp mindestens 25 Jahre an Daten zur Verfügung stehen müssen. Dies ist insofern kritisch, da sich somit der betrachtete Länderumfang stark reduziert<sup>28</sup>, obwohl es in der Praxis per se keinen Grund gibt, aus dem sich ein hypothetischer Investor selbst in seinen Opportunitäten einschränken sollte. Auch die definierte Aufspaltung des Gesamtzeitraums in eine statische In- und Out-of-Sample-Periode wirkt in diesem Zusammenhang zu einem gewissen Grad willkürlich.

Als zweiter Kritikpunkt kann die fehlende Berücksichtigung von Transaktionskosten, etwa in der Form von Bid- und Ask-Spreads, angeführt werden. Diese unrealistische Betrachtung führt im Normalfall zu einer chronischen Überschätzung der Performance. Da aber ohnehin keines der optimierten Portfolios die starke In-Sample-Performance im Out-of-Sample-Test beibehalten konnte, wurde bis auf Weiteres von der Berücksichtigung von Transaktionskosten abgesehen. Außerdem ist diskutabel, ob die Abstinenz jeglicher Leverage-Restriktionen eine sinnvolle Annahme für einen durchschnittlichen Investor ist.

Ebenfalls Anlass zur Kritik kann womöglich die genaue methodologische Definition der verwendeten Charakteristika bieten. So könnte zur Berechnung des realen Wechselkurses der Produzentenpreis- statt dem Konsumentenpreisindex Verwendung finden. Dem folgend wäre etwa auch die Frage interessant, ob ein Signal basierend auf dem realen effektiven Wechselkurs, welcher in der Praxis mitunter als Bewertungsmaßstab für Währungen herangezogen wird, zu anderen Ergebnissen führen würde. Ähnliche Fragestellungen können auch für alle anderen Charakteristika aufgeworfen werden. Im Detail kann zudem auch der definierte Umgang mit der zeitverzögerten Veröffentlichung von Daten kritisiert werden, welcher unabhängig vom Datentyp pauschal eine Verzögerung von zwei Monaten unterstellt.

Der letzte Kritikpunkt dreht sich um die Tatsache, dass die Zusammensetzung des optimalen Portfolios beim verwendeten Verfahren maßgeblich von den unterstellten Präferenzen des Investors abhängt. In der bisherigen Analyse habe ich in Anlehnung an Barroso und Santa-Clara (2015) die in Abschnitt 2.2 vorgestellte CRRA-Nutzenfunktion mit  $\gamma=4$  verwendet. Daher ist es sicherlich interessant zu erfahren, wie sich die Ergebnisse bei einer Änderung des Grads der relativen Risikoaversion oder gar bei einer anderen Nutzenfunktion darstellen. Im Zuge der nächsten Kapitel werden einige ausgewählte der genannten Kritikpunkte behandelt, wobei

<sup>27</sup>Zur Erinnerung: Der Datensatz von Barroso und Santa-Clara (2015) beinhaltet maximal 21 und minimal 13 Währungen, während der hier untersuchte Datensatz maximal 18 und minimal zehn Währungen abdeckt.

<sup>28</sup>Hier sollte daran erinnert werden, dass sich im untersuchten Datensatz keine einzige als Entwicklungsland klassifizierte Nation befand.



eine Abdeckung aller gegebenen des Umfangs der vorliegenden Arbeit nicht möglich ist.

## 6. Variation des Grads der relativen Risikoaversion

Die bisherige Analyse basierte, inspiriert von Barroso und Santa-Clara (2015) und wie in Abschnitt 2.2 dargestellt, auf einem Aversionsparameter der CRRA-Nutzenfunktion von  $\gamma=4$ . Ähnlich zu Brandt et al. (2009) habe ich als Modellvalidierung<sup>29</sup> in einer Szenarioanalyse betrachtet, wie sich die wesentlichen In-Sample-Ergebnisse aus Abschnitt 5.2 ändern, wenn der Grad der relativen Risikoaversion ab- bzw. zunimmt. Hierzu wurde eine Halbierung ( $\gamma=2$ ) sowie ein extremer Anstieg ( $\gamma=100$ ) als Alternativszenarien untersucht. Die Ergebnisse der Auswertung in Tabellenformat finden sich in Anhang E.

Brandt et al. (2009) beobachten bspw. im Zuge ihrer vergleichbaren Analyse, dass bei Erhöhung der relativen Risikoaversion zwei ihrer drei Charakteristika-Koeffizienten null entgegenstreben, während der verbliebene stabil bleibt. Sie schließen hieraus, dass die besagten beiden Charakteristika mit sowohl Rendite wie auch Risiko in Zusammenhang stehen und der Investor mit der CRRA-Nutzenfunktion bei steigendem Grad der relativen Risikoaversion der Risikokontribution mehr Beachtung schenkt, während der Koeffizient der verbliebenen Charakteristik eher mit erwarteten Renditen denn Risiko assoziiert werden kann. In der vorliegenden Arbeit konnte ich für ausnahmslos alle Charakteristika im Zuge jeder untersuchten Strategie eine Annäherung zum Nullwert bei steigendem Aversionsparameter feststellen. Die prozentuale Änderung der Koeffizienten über die drei Szenarien gestaltet sich dabei recht uniform – sowohl beim Blick auf eine spezifische Charakteristik als auch über alle dargestellten Charakteristika hinweg. Betrachtet man die durchschnittliche Änderung des Koeffizienten einer Charakteristik im Kontext aller untersuchten Strategien, die jene spezifische Charakteristik beinhalten, so nahm etwa beim Sprung von  $\gamma=2$  auf  $\gamma=4$  der Koeffizient für Carry um -46,7% ab, mit einer geschätzten Standardabweichung von lediglich 0,8%. Für die anderen Charakteristika sowie den Sprung von  $\gamma=4$  auf  $\gamma=100$  sind die entsprechenden Werte in folgender Tabelle abgetragen:

Die Tabelle zusammenfassend näherten sich die Koeffizienten aller Charakteristika bei der Verdopplung von  $\gamma=2$  auf  $\gamma=4$  im Schnitt mit einer Veränderung von knapp 46,2% null an und wiesen dabei eine Standardabweichung von 3% auf,

während bei der Erhöhung von  $\gamma=4$  auf  $\gamma=100$  eine durchschnittliche Änderung von 95,7% mit einer Standardabweichung von 0,8% beobachtet werden konnte. Der Argumentation von Brandt et al. (2009) folgend scheint es also, dass die selektierten Charakteristika – zumindest bei Betrachtung der in Abschnitt 4.3 präsentierten Datenauswahl – durchaus mit Risiko in Zusammenhang stehen und der CRRA-Investor bei steigender Risikoaversion daher weniger Exposure zu diesen präferiert. Zu erwähnen bleibt, dass trotz der Abnahme der geschätzten Koeffizienten bei den Signifikanzniveaus keine nennenswerten Änderungen zu beobachten waren.

Betrachtet man außerdem die Entwicklung der verschiedenen Performance-Kennzahlen der In-Sample-Periode bei Variation der relativen Risikoaversion, zu finden in Tabelle E.22 des Anhangs E, können ähnliche Beobachtungen gemacht werden, wie auch schon Brandt et al. (2009) sie tätigten. So sind wenig überraschend die höchsten durchschnittlichen und maximalen Renditen sowie Volatilitäten wie auch die negativsten minimalen Renditen bei den optimierten Strategien mit  $\gamma=2$  festzustellen; bei steigender Risikoaversion nähern sich alle Kennzahlen null an. Grund hierfür ist, wie Brandt et al. (2009) ausführen, dass der durchschnittliche CRRA-Nutzen über alle Monate der Betrachtungsperiode bei einem sehr hohen Aversionsparameter maßgeblich vom Nutzen im schlechtesten Monat beeinträchtigt wird und somit einem Minimum/Maximum-Kriterium gleicht. Bemerkenswert ist in diesem Zusammenhang der Fakt, dass die In-Sample-Optimierung mit einem Aversionsparameter von  $\gamma=100$  für die Einzelstrategien auf Basis des Leistungsbilanzsaldos sowie Volatilitätsfaktors zwei Portfolios generieren konnte, welche eine positive minimale Rendite aufweisen.

Ebenfalls vergleichbar zu Brandt et al. (2009) sowie intuitiv nachvollziehbar agiert der risikofreudigste Investor stets mit dem höchsten Leverage. Bei steigender relativer Risikoaversion sinkt der Leverage schließlich für sämtliche betrachteten Strategien. Zudem nimmt bei Anstieg des Aversionsparameters auch stets der Umschlag des Forward-Portfolios ab, was ebenso in Einklang mit den Ergebnissen von Brandt et al. (2009) ist. Hier sei als Kuriosität auf die Einzelstrategie des Leistungsbilanzsaldos verwiesen, welche bei  $\gamma=100$  im Schnitt einen auf zwei Nachkommastellen gerundeten Umschlag von null auswies. Anders ausgedrückt glich sie damit im Durchschnitt einer Strategie, die jeden Monat das immer wieder gleiche Portfolio aus Forward-Währungskontrakten mit derselben Gewichtung hielt. Wesentlich zurückzuführen ist dies auf den geschätzten Koeffizienten von 0,02, welcher sich von allen untersuchten Strategien am meisten null angenähert hat, wobei auch bspw. die im Vergleich zu anderen Datentypen relative niedrige Frequenz des Leistungsbilanzsaldos eine Rolle gespielt haben mag.

Außerdem sei auf ein Ergebnis aufmerksam gemacht, dass ich im Zuge der Korrelationsanalyse, zu finden in Tabelle E.24 des Anhangs E, feststellen konnte. So führte der Anstieg von  $\gamma=2$  auf  $\gamma=4$  zwar zu keiner wesentlichen Änderung der Korrelationen zwischen den Einzelstrategien, der folgende massive Anstieg von  $\gamma=4$  auf  $\gamma=100$  dafür umso

<sup>29</sup>Obschon ein direkter Vergleich zu Brandt et al. (2009) nicht möglich ist, da die Autoren das Verfahren im Aktienkontext anwenden, lässt sich ein indirekter Vergleich ziehen. So betrachten die Autoren ähnlich zur vorliegenden Arbeit eine CRRA-Nutzenfunktion als Zielgröße der Optimierung. Es ist daher anzunehmen, dass mit einer vergleichbaren Variation des Aversionsparameters eine ähnliche Veränderung der Portfolio-Eigenschaften einhergehen sollte, da die Präferenz hinsichtlich dieser Eigenschaften – etwa durchschnittliche Rendite oder Standardabweichung – nur durch die CRRA-Nutzenfunktion definiert wird und damit natürlich unabhängig von bspw. der Anlageklasse ist.

**Tabelle 5:** Änderung der geschätzten Koeffizienten bei variierender relativer Risikoaversion

Die Tabelle zeigt Durchschnitt ( $\bar{\Delta}$ ) und Standardabweichung ( $\sigma(\Delta)$ ) der prozentualen Änderung des Koeffizienten einer Charakteristik bei variierender relativer Risikoaversion über alle untersuchten Strategien hinweg, die jene spezifische Charakteristik beinhalten. Sie basiert auf der in Anhang E zu findenden Tabelle E.23, welche – jeweils für die unterschiedlichen Aversionsparameter  $\gamma = 2$ ,  $\gamma = 4$  und  $\gamma = 100$  – die über den Zeitraum der In-Sample-Periode von April 1990 bis Oktober 2005 geschätzten Koeffizienten aller optimierten Strategien darstellt. Demnach wird die Erhöhung der Aversion von  $\gamma = 2 \rightarrow 4$  und von  $\gamma = 4 \rightarrow 100$  betrachtet. Namentlich handelt es sich bei den Charakteristika um folgende: Carry ( $fd$ ), Momentum ( $mom$ ), langfristige Trendumkehr ( $q$ ), Leistungsbilanzsaldo ( $ca$ ), die Produktionslücke ( $opg$ ) sowie der Volatilitätsfaktor ( $vol$ ). Die Ergebnisse wurden auf eine Nachkommastelle gerundet.

|                              |                  | $fd$    | $mom$   | $q$     | $ca$    | $opg$   | $vol$   |
|------------------------------|------------------|---------|---------|---------|---------|---------|---------|
| $\gamma = 2 \rightarrow 4$   | $\bar{\Delta}$   | -46,70% | -50,20% | -46,30% | -47,90% | -44,70% | -41,30% |
|                              | $\sigma(\Delta)$ | 0,80%   | 0,80%   | 1,00%   | 2,20%   | 12,10%  | 7,90%   |
| $\gamma = 4 \rightarrow 100$ | $\bar{\Delta}$   | -96,10% | -96,20% | -95,70% | -96,40% | -94,20% | -95,90% |
|                              | $\sigma(\Delta)$ | 0,10%   | 0,20%   | 0,10%   | 0,30%   | 1,00%   | 0,20%   |

mehr. Demnach sind in Folge die Korrelationskoeffizienten zwischen ausnahmslos allen Einzelstrategien angestiegen. Definiert man weiter eine hohe Korrelation arbiträr als  $>0,7$ , so konnte bei den beiden Szenarien mit vergleichsweise niedriger relativer Risikoaversion nur die Korrelation zwischen Carry-Einzelstrategie und -Benchmark (0,96) als hoch bezeichnet werden – und dies aus nachvollziehbarem Grund. Im Zuge des Szenarios mit  $\gamma=100$  würden nun sieben Korrelationskoeffizienten in diese Kategorie fallen. Besonders ins Auge springen dabei die hohen Korrelationen, die nun im Allgemeinen zwischen Momentum, dem Leistungsbilanzsaldo, der Produktionslücke und dem Volatilitätsfaktor bestehen. Es scheint, dass sich die von einem sehr risikoaversen Investor optimierten Strategien auf Basis dieser vier Signale nicht mehr verhältnismäßig stark in ihren Renditebewegungen zu unterscheiden scheinen und somit aus Diversifikationsaspekten weniger interessant sind. Dies wird bspw. auch beim Blick auf die Sharpe Ratios im entsprechenden Risikoszenario deutlich. So bringen die genannten vier Charakteristika relativ weniger Mehrwert bei der Inklusion im Zuge der Kombinationsstrategien. In Anlehnung an [Brandt et al. \(2009\)](#) wage ich die Deutung, dass die Renditen dieser Einzelstrategie für einen sehr risikoaversen Investor als zu riskant erscheinen, um sie mit Mehrwert als Steuerungssignal verwenden zu können.

Zuletzt habe ich untersucht, ob die aufgeführten Beobachtungen hinsichtlich der Performance-Kennzahlen auch in der Out-of-Sample-Analyse gemacht werden können. Aufschluss über diese Analyse gibt Tabelle E.25 in Anhang E. Im Wesentlichen stimmen die Ergebnisse mit den In-Sample-Resultaten überein. Lediglich bei den durchschnittlichen Renditen finden sich für Strategien bei höherem Aversionsparameter im Gegensatz zu den vorigen Beobachtungen teilweise höhere Renditen und nun kann auch keine untersuchte Strategie mehr eine positive minimale Rendite aufweisen.

Vergleicht man die Sharpe Ratios im Out-of-Sample-Test bei variierender Risikoaversion, befinden sich diese für die jeweiligen Strategien bei alle drei Szenarien in ähnlichen Regionen – wie zuvor auch im In-Sample-Test. Das Optimierungsverfahren skaliert demnach in Einklang mit der gewählten Nutzenfunktion Mittelwert und Standardabweichung in etwa linear nach unten in Anbetracht sinkender Risikotoleranz. Beim Sicherheitsäquivalent sanken innerhalb des In-

Sample-Tests die Werte bei steigender Risikoaversion oder hielten zumindest ihr Niveau – anderes lässt sich nun bei den Out-of-Sample-Ergebnissen feststellen. Konnten beim Standardszenario mit  $\gamma=4$  fast alle optimierten Strategien nur ein negatives Sicherheitsäquivalent erreichen und erschienen neben der US-Aktien-Benchmark chancenlos, so hat sich das Resultat aus der Perspektive eines extrem risikoaversen Investors geändert. Diesem bieten die optimierten Portfolios, welche die Informationen der Charakteristika mit vergleichsweise geringem Leverage und Umschlag in den Portfoliogewichten verarbeiten, zumindest leicht positive Sicherheitsäquivalente. Eine Investition in den Aktienmarkt erscheint einem solchen Investor dagegen unvorteilhaft.

## 7. Expandierendes Währungsuniversum

### 7.1. Änderungen in Methodik und Datensatz

Wie in Abschnitt 5.4 ausgeführt, ist der bisher betrachtete Länder- bzw. Währungsumfang, welcher sich aus einer geforderten Mindestanzahl von 25 Jahren pro Datentyp und Währung ergab, vergleichsweise gering und ein möglicher Kritikpunkt am Vorgehen. Um einen größeren Nutzen aus den akquirierten und in Abschnitt 4.1 sowie Anhang A ersichtlichen Daten zu ziehen, gehe ich in der weiteren Analyse wie folgt vor: Der neue Datensatz erstreckt sich zwar über die gleiche Zeitspanne von Ende April 1990 bis Ende Dezember 2017, wie in der bisherigen Analyse, allerdings beinhaltet er nun alle Währungen, für die zu jedem nötigen Datentyp eine Zeitreihe vorliegt – zunächst unabhängig davon, in welcher zeitlichen Länge die Reihen zur Verfügung stehen. Von den 42 Ländern des angestrebten Datensatzes aus Abschnitt 4.1 bleiben somit 28 Länder<sup>30</sup> übrig.

Der Unterschied in der Methodik ist die Definition des zu jedem Zeitpunkt investierbaren Währungsuniversums  $N_t$ . Bestand bisher ein statisches Universum aus Währungen mit der geforderten Mindestanzahl von 25 Jahren pro

<sup>30</sup>Auf welche Datenquelle für die verschiedenen Datentypen je Land zurückgegriffen wurde, kann in Anhang C nachvollzogen werden. Die Slowakei wurde ausgeschlossen, da für sie nur an einem einzigen Monat vor der nationalen Euro-Einführung alle Daten zur Verfügung standen und daher eine Exklusion keinen wesentlichen Unterschied macht.

Datentyp, welches nur punktuell durch die definierten Euro-Einführungsdaten beeinflusst wurde, wird nun ein komplexeres Vorgehen gewählt. So wird zu jedem Zeitpunkt  $t$  für alle 28 Länder geprüft, ob die nötigen Angaben zur Berechnung aller Charakteristika vorliegen. Ist dies erstmals der Fall, wird die entsprechende Währung fortan in das Universum inkludiert. Im Ergebnis finden sich nun minimal 16 und maximal 26 Währungen zeitgleich im Datensatz, während es bei der bisherigen statischen Analyse nur minimal zehn und maximal 18 und bei Barroso und Santa-Clara (2015) minimal 13 und maximal 21 waren. Abbildung 3 zeigt in diesem Zusammenhang, wie sich das Währungsuniversum über die Gesamtperiode verändert – sowohl für die statische Definition aus den vorherigen Kapiteln als auch das expandierende Währungsuniversum, welches im vorliegenden Kapitel betrachtet wird und zusätzlich vier weitere Industrienationen und fünf Entwicklungsstaaten<sup>31</sup> miteinschließt.

Um Vergleichbarkeit mit den bisherigen Ergebnissen zu ermöglichen, wird die sonstige Methodik, wie etwa das Vorgehen bei spezifischen Fragestellungen zur Datenverwendung aus Abschnitt 4.2 oder die in Abschnitt 5.1 beschriebene Aufteilung des Datensatzes in In- und Out-of-Sample-Periode, nicht angetastet.

## 7.2. In-Sample-Ergebnisse

### 7.2.1. Performance

Tabelle 6 präsentiert die Performance der optimierten Portfolios sowie jene der definierten Benchmarks über die Zeitperiode von April 1990 bis Oktober 2005 bei Verwendung des expandierenden Währungsuniversums – wie zuvor Tabelle 2 in Abschnitt 5.2.1 für das statische Universum. Zusammenfassend kann festgestellt werden, dass sich die erweiterte Definition des Währungsuniversums massiv auf die abgebildeten Kennzahlen der In-Sample-Periode ausgewirkt hat.

Bei den Einzelstrategien sind durchschnittliche Rendite und Standardabweichung im Vergleich zu den vorigen In-Sample-Ergebnissen im Mittel um 171% bzw. 190% angestiegen. Beides resultiert in einem durchschnittlichen Anstieg der Sharpe Ratios von 117%. Auch minimale und maximale Renditen der Einzelstrategien verzeichnen fast ausnahmslos extremere Werte als bei Verwendung des statischen Währungsuniversums, zudem ist im Schnitt eine deutlich höhere Exzess-Kurtosis zu beobachten. Vier statt zuvor zwei der sechs Einzelstrategien zeigen nun ein positives Vorzeichen bei der Schiefe und auch das Sicherheitsäquivalent steigt im Durchschnitt um 92%.

Eine wesentliche Ausnahme zu diesen Durchschnittsbeobachtungen ist die Einzelstrategie auf Basis der langfristigen Trendumkehr, welche sich gänzlich anders im Verhältnis zum vorigen In-Sample-Test entwickelt hat. Die durchschnittliche Rendite und Standardabweichung der Strategie

ist um 76% bzw. 77% gesunken und auch minimale sowie maximale Rendite deutlicher näher in Richtung null gerückt, während sich die Exzess-Kurtosis ebenfalls um 66% verringert hat. Zudem weist die Strategie nun eine negative Schiefe auf. In der Konsequenz handelt es sich bei der langfristigen Trendumkehr um die einzige Einzelstrategie, welche hinsichtlich Sharpe Ratio und Sicherheitsäquivalent Einbußen im Vergleich zum vorher untersuchten Währungsuniversum hinnehmen musste. Sollen hingegen bzgl. beider risikoadjustierter Performance-Größen die Einzelstrategien mit den höchsten Zuwächsen im Vergleich zur Analyse aus Abschnitt 5.2.1 benannt werden, handelt es sich mit Blick auf die Sharpe Ratio um den Leistungsbilanzsaldo mit einem Anstieg von 350% und Momentum mit 222%. Beim Sicherheitsäquivalent weisen Momentum (321%) und Carry (227%) den höchsten Anstieg auf, was u.a. mit der nun stark positiven Schiefe bei der Einzelstrategien zusammenhängt.

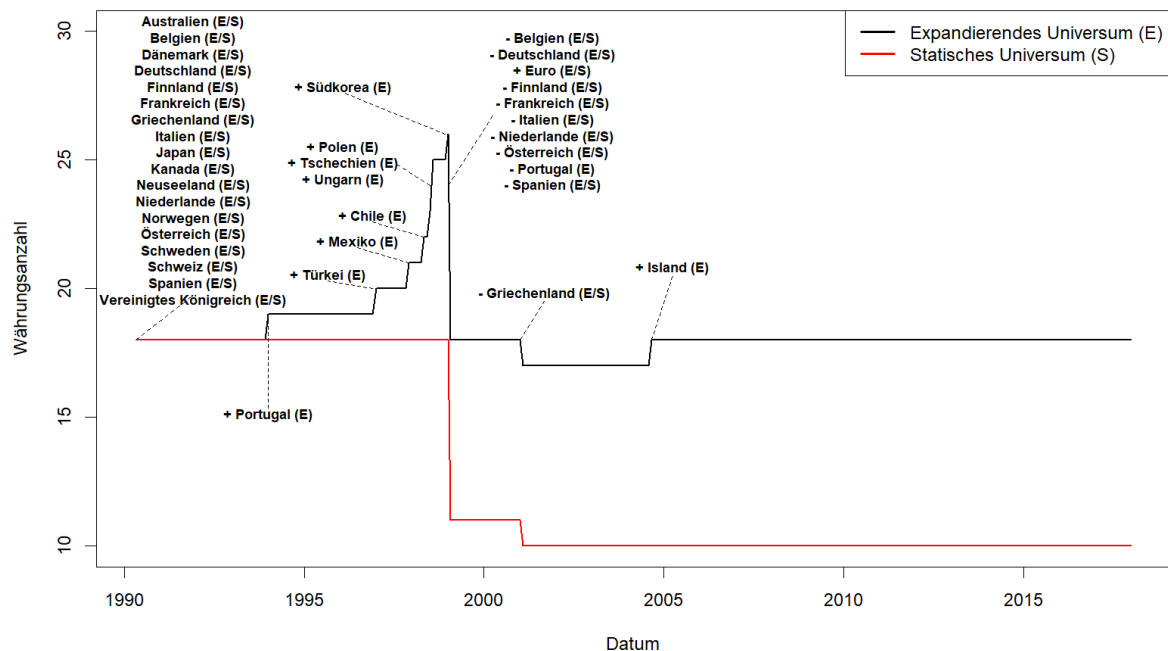
Um ein besseres Verständnis zu erlangen, wie die Performance der optimierten Einzelstrategien in Zusammenhang steht, habe ich wie in den vorigen Kapiteln eine Korrelationsanalyse durchgeführt, die in Tabelle F.26 des Anhangs F dargestellt ist. Es zeigt sich, dass Momentum und Produktionslücke, die beiden Einzelstrategien mit den höchsten Sharpe Ratios (0,29 und 0,12) nach Carry (0,47), vergleichsweise stark mit der Carry-Einzelstrategie korreliert sind (0,7 und 0,6). Die Leistungsbilanzsaldo-Einzelstrategie, deren Sharpe Ratio (0,09) sich auf Rang vier unter den Einzelstrategien befindet, hat dagegen nur eine Korrelation von 0,18 mit der Carry-Einzelstrategie. Bemerkenswert ist hier außerdem, dass die langfristige Trendumkehr als Einzelstrategie zwar die geringste Sharpe Ratio (0,05) erzielen kann, dafür aber mit allen Einzelstrategien eine negative Korrelation aufweist. Zuletzt sei darauf verwiesen, dass zwischen der US-Aktien-Benchmark und den Einzelstrategien im Schnitt nur eine Korrelation von 0,02 vorliegt.

Im Zuge der Kombinationsstrategien können im Vergleich zum statischen Währungsuniversums ähnliche Beobachtungen wie für die Einzelstrategien festgestellt werden. Demnach sind vor allem maximale und durchschnittliche Rendite, Standardabweichung und Exzess-Kurtosis massiv angestiegen und die Schiefe nun für alle Strategien stark positiv. Allerdings kann bei Betrachtung der Sharpe Ratios aller sukzessiven Kombinationsstrategien nur ein durchschnittlicher Anstieg von 8% konstatiert werden, wohingegen das Sicherheitsäquivalent nun im Schnitt um 234% höher ist. Zudem sind auch Leverage und Umschlag bei allen sukzessiven Kombinationsstrategien im Mittel um 60% bzw. 36% angestiegen.

### 7.2.2. Koeffizienten und statistische Signifikanz

In Tabelle 7 sind zu jeder Strategie die geschätzten Koeffizienten sowie jeweils in Klammern darunter deren Signifikanz in Form von p-Werten abgebildet – wie zuvor in Tabelle 3 des Abschnitts 5.2.2 für das statische Universum. Das Vorzeichen der Produktionslücke-Einzelstrategie und jene der Volatilität-Kombinationsstrategien entsprechen nicht den Erwartungen.

<sup>31</sup>Zur Erinnerung: Die Einteilung in Industrie- und Entwicklungsländer erfolgt wie in Abschnitt 4.1 beschrieben in Übereinstimmung mit dem IWF (IWF (2018)).



**Abbildung 3:** Expandierendes vs. statisches Währungsuniversum.

Die Abbildung stellt die Anzahl der investierbaren Währungen über den Zeitablauf der Gesamtperiode von Ende April 1990 bis Ende des Jahres 2017 für die zwei verwendeten Definitionsansätze des Währungsuniversums dar. In roter Farbe abgetragen ist das statische Vorgehen, welches in den vorherigen Kapiteln Anwendung fand und initial auf der Voraussetzung basiert, dass pro Land und Datentyp mindestens 25 Jahre an Daten zur Verfügung stehen müssen. Es wird nur durch die in Abschnitt 4.2 genannten Euro-Einführungsdaten beeinflusst. In schwarzer Farbe abgetragen ist das expandierende Währungsuniversum, welches im Zuge des vorliegenden Kapitels analysiert wird. Es beinhaltet zu jedem Zeitpunkt alle Währungen, für welche alle in Abschnitt 4.1 aufgeführten Datentypen zur Verfügung stehen. Am Ende einer gestrichelten Linie werden bei einer Änderung des Währungsuniversums immer jene Länder genannt, welche die jeweilige Änderung verursacht haben. Dabei besteht die Möglichkeit, dass eine Währung bereits initial im Datensatz vorhanden ist (kein Zeichen vor dem Namen) oder in das Universum inkludiert (+ vor dem Namen) bzw. im Zuge der Euro-Einführung exkludiert wird (- vor dem Namen). Hinter dem Ländernamen ist in Klammern angegeben, ob der expandierende Ansatz (E) und/oder der statische Ansatz (S) von der Änderung betroffen ist.

Bei Betrachtung der Koeffizienten der Einzelstrategien zeichnen sich auf dem 5%- sowie 1%-Signifikanzniveau von allen Charakteristika nun Carry und Momentum als relevant ab, während dies beim statischen Währungsuniversum für Carry und die langfristige Trendumkehr galt. Die Charakteristik auf Basis des realen Wechselkurses zeigt aber in allen untersuchten Strategiekombinationen Signifikanz auf dem 1%-Niveau auf. Gleiches gilt für den Leistungsbilanzsaldo.

Die stärkste prozentuale Änderung im Vergleich zum statischen Währungsuniversum ist bei den Koeffizienten der Leistungsbilanz-Charakteristik zu beobachten, welche im Schnitt um fast 190% anstiegen. Danach folgt jene von Momentum mit einem durchschnittlichen Anstieg von ca. 117% und Carry mit etwa 71%. In Folge weist Carry nun im Schnitt den höchsten Koeffizienten auf (21,32), gefolgt von Momentum (11,12), dem Leistungsbilanzsaldo (10,66) und der langfristigen Trendumkehr (9,19). Die durchschnittlichen Koeffizienten von Produktionslücke (1,33) und Volatilitätsfaktor (-1,63) sind dagegen deutlicher näher bei null und erscheinen demnach weniger statistisch signifikant.

Ein interessantes Ergebnis zeigt sich mit Blick auf die Änderungen der Koeffizienten im Zuge der Kombinationsstrate-

gien. Während sich die Koeffizienten der Carry-Momentum-Kombinationsstrategie beim statischen In-Sample-Test nicht wesentlich zu denen der beiden entsprechenden Einzelstrategien unterscheiden, kann jetzt ein Anstieg verzeichnet werden. Ähnliches gilt beim Hinzufügen der langfristigen Trendumkehr. Die Koeffizienten der drei Charakteristika steigen erneut, wobei für Carry und Momentum eine moderate Vergrößerung im Vergleich zur Kombinationsstrategie ohne langfristige Trendumkehr zu verzeichnen ist, während der Koeffizient der langfristigen Trendumkehr relativ stark zur entsprechenden Einzelstrategie ansteigt. Nochmals kann dieses Verhalten im Zuge der Inklusion des Leistungsbilanzsaldos betrachtet werden. Bei Inklusion der insignifikanten Produktionslücke kann in Folge keine wesentliche Veränderung der Koeffizienten betrachtet werden, während das Einbeziehen des auf dem 5%-Niveau signifikanten Volatilitätsfaktors zu einer Verminderung der Koeffizienten von Momentum und langfristiger Trendumkehr führt, während jene von Carry und Leistungsbilanzsaldo steigen.

Auf nähere Erläuterungen hinsichtlich der Veränderung der geschätzten Koeffizienten über den Zeitablauf der In-Sample-Periode, welche zuvor in Abbildung 1 des Abschnitts



**Tabelle 6:** In-Sample-Performance bei expandierendem Währungsuniversum

Die Tabelle stellt die Performance-Kennzahlen der optimierten Strategien sowie der Benchmarks über den Zeitraum der In-Sample-Periode von April 1990 bis Oktober 2005 bei Verwendung des expandierenden Währungsuniversums dar. Die Optimierung betrachtet eine CRRA-Nutzenfunktion mit  $\gamma = 4$ . Aufgeführt werden in den Spalten die maximale ( $\max r$ ), minimale ( $\min r$ ) sowie durchschnittliche ( $\bar{r}$ ) monatliche Rendite, die geschätzte monatliche Standardabweichung ( $\sigma(r)$ ) (alle bisherigen Angaben in Prozent), Exzess-Kurtosis und Schiefe der Wahrscheinlichkeitsverteilung, die monatliche Sharpe Ratio (SR) und das monatliche Sicherheitsäquivalent in Prozent (SE). Die letzten beiden Spalten geben Auskunft über den durchschnittlichen Leverage ( $\bar{L}$ ), angegeben als der Betrag, welcher pro in die risikolose Geldanlage investiertem Dollar in das riskante Währungsportfolio angelegt ist, und Umschlag ( $\bar{T}$ ), der sich als die Summe aus den absoluten Änderungen aller Währungsgewichte des Forward-Portfolios zwischen zwei Perioden definiert. In den ersten sechs Zeilen werden die Kennzahlen für die Einzelstrategien abgebildet. Eine Einzelstrategie bezeichnet dabei das Resultat einer Optimierung, welche nur mit einer einzelnen Charakteristik als Inputfaktor arbeitet. Namentlich sind das folgende Strategien: Carry (fd), Momentum (mom), langfristige Trendumkehr (q), Leistungsbilanzsaldo (ca), die Produktionslücke (opg) sowie der Volatilitätsfaktor (vol). In den nächsten fünf Zeilen folgt die Betrachtung von Strategiekombinationen. Hierfür wurde der Optimierung in einer arbiträren Reihenfolge sukzessiv immer eine weitere Charakteristik als Inputfaktor hinzugefügt, bis die finale Strategie alle vorhandenen Charakteristika miteinbezieht. Wegen der arbiträren Reihenfolge finden sich in den folgenden zwei Zeilen zwei zusätzliche Strategiekombinationen, welche die drei aus der vorangegangenen Analyse relevantesten Charakteristika Carry, Momentum und langfristige Trendumkehr mit jeweils der Produktionslücke und dem Volatilitätsfaktor paaren. In den letzten drei Zeilen der Tabelle findet sich die Performance der drei Benchmarks. Für die Berechnung des Umschlages der US-Aktien Benchmark lagen keine Daten vor.

| Strategie     | $\max r$ | $\min r$ | $\bar{r}$ | $\sigma(r)$ | Kurtosis | Schiefe | SR   | SE    | $\bar{L}$ | $\bar{T}$ |
|---------------|----------|----------|-----------|-------------|----------|---------|------|-------|-----------|-----------|
| 1. <i>fd</i>  | 146.51   | -32.28   | 14.16     | 29.71       | 7.83     | 2.64    | 0.47 | 5.79  | 9.59      | 7.43      |
| 2. <i>mom</i> | 54.81    | -21.51   | 3.72      | 11.55       | 6.13     | 2.20    | 0.29 | 1.81  | 5.21      | 3.84      |
| 3. <i>q</i>   | 3.81     | -3.78    | 0.40      | 1.32        | 0.50     | -0.27   | 0.05 | 0.37  | 1.21      | 0.26      |
| 4. <i>ca</i>  | 9.57     | -4.71    | 0.56      | 2.42        | 0.50     | 0.45    | 0.09 | 0.44  | 2.72      | 0.18      |
| 5. <i>opg</i> | 16.29    | -11.79   | 0.71      | 3.15        | 7.16     | 0.84    | 0.12 | 0.52  | 2.06      | 0.15      |
| 6. <i>vol</i> | 4.43     | -7.32    | 0.41      | 1.39        | 4.44     | -0.63   | 0.06 | 0.37  | 0.91      | 0.79      |
| 1 - 2         | 247.62   | -26.30   | 22.36     | 49.59       | 9.57     | 3.07    | 0.44 | 7.36  | 14.34     | 11.67     |
| 1 - 3         | 266.19   | -36.79   | 25.47     | 53.40       | 9.28     | 3.04    | 0.47 | 8.70  | 16.60     | 12.91     |
| 1 - 4         | 324.25   | -34.86   | 30.72     | 63.18       | 9.21     | 3.03    | 0.48 | 9.91  | 22.25     | 14.83     |
| 1 - 5         | 318.30   | -36.12   | 30.48     | 61.86       | 8.94     | 2.99    | 0.49 | 9.97  | 22.09     | 14.85     |
| 1 - 6         | 320.20   | -37.58   | 30.83     | 61.92       | 9.12     | 3.01    | 0.49 | 10.28 | 22.82     | 15.60     |
| 1 - 3 & 5     | 260.11   | -38.94   | 25.43     | 51.97       | 8.83     | 2.97    | 0.48 | 8.92  | 16.78     | 13.17     |
| 1 - 3 & 6     | 266.18   | -37.08   | 25.47     | 53.39       | 9.34     | 3.04    | 0.47 | 8.71  | 16.57     | 12.96     |
| Carry         | 227.22   | -41.51   | 18.65     | 41.63       | 10.25    | 3.05    | 0.44 | 5.67  | 11.90     | 9.22      |
| Naiv          | 5.68     | -7.32    | 0.36      | 2.27        | 0.27     | -0.24   | 0.01 | 0.26  | 1.00      | 0.01      |
| US-Aktien     | 11.22    | -15.65   | 0.96      | 4.27        | 0.74     | -0.58   | 0.15 | 0.58  | 0.00      | -         |

5.2.2 für das statische Universum präsentiert wurde, wird hier verzichtet, da sich die Veränderung nicht wesentlich zur vorigen Analyse unterscheidet, auch wenn sich die Koeffizienten i.d.R. nun auf höheren Wertniveaus einpendeln. Zur Transparenz findet sich die entsprechende Abbildung F.7 in Anhang E.

Zusätzlich habe ich – um eine bessere Aussage über die Stabilität der Koeffizienten des Verfahrens über den Zeitablauf treffen zu können – für das expandierende Währungsuniversum untersucht, wie sich die geschätzten Koeffizienten bei monatlicher Schätzung mit einem gleitenden statt des sonst verwendeten expandierenden Datenfensters verändern. Abbildung 4 präsentiert die Ergebnisse dieser Analyse bei Verwendung eines gleitenden Datenfensters von fünf Jahren. Es wird deutlich, dass der Einfluss der Charakteristika auf die Portfoliogewichte über die In-Sample-Periode starken Änderungen unterworfen ist. Zudem kann etwa anhand der Entwicklung des Carry-Koeffizienten, aber auch anderer Charakteristika beobachtet werden, wie das In- und Exkludieren von Währungen zu einer Art Regimewechsel führt. Zu Beginn des Jahres 1999, also gemäß Abbildung 3 kurz nachdem der Großteil der zusätzlichen Währungen einbezogen wurden und zu dem Zeitpunkt, an welchem im Zuge der Euro-

Einführung verschiedene Währungen das Universum verlassen, ist eine Änderung in den Koeffizienten zu betrachten, jeweils in unterschiedlich starker Ausprägung. Für Carry, Leistungsbilanzsaldo und langfristige Trendumkehr ist bspw. ein plötzlicher Anstieg der Koeffizienten die Folge. Da allerdings die Inklusion der zusätzlichen Währungen zeitlich sehr nahe mit der Exklusion der Mehrzahl von Euro-Länder zusammenliegt, kann nicht zweifelsfrei bestimmt werden, was der ausschlaggebende Faktor für diese Beobachtung ist. Anfang des Jahres 2004, etwa fünf Jahre später und somit zu jenem Zeitpunkt, an dem das gleitende Datenfenster die Daten zum Jahresbeginn 1999 nicht länger inkludiert, sind weitere extreme Änderungen – z.B. ein massiver Einbruch des Carry-Koeffizienten – zu beobachten.

Tabelle 8 liefert eine einfache quantitative Analyse der in Abbildung 4 präsentierten Daten. Demnach bewegt sich der Leistungsbilanz-Koeffizient innerhalb der größten Schwankungsbreite (98,98), gefolgt von der Produktionslücke (80,63) und an dritter Stelle Carry (63,52), aber auch die Schwankungsbreite der anderen Charakteristika-Koeffizienten ist substantiell.

Wirft man einen Blick auf die Mittelwerte der Koeffizienten, so hat Carry im Durchschnitt mit Abstand den größten



**Tabelle 7:** Geschätzte Koeffizienten und Signifikanz bei expandierendem Währungsuniversum

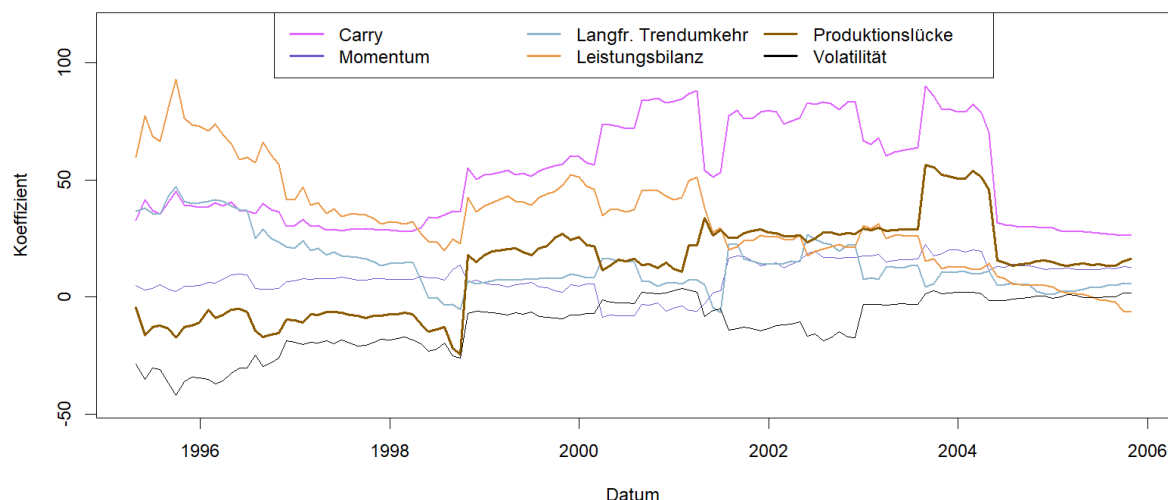
Die Tabelle zeigt für jede optimierte Strategie die über den Zeitraum der In-Sample-Periode von April 1990 bis Oktober 2005 bei Verwendung des expandierenden Währungsuniversums geschätzten Koeffizienten der Charakteristika sowie jeweils in Klammern darunter deren Signifikanz in Form von p-Werten. Namentlich handelt es sich bei den Charakteristika um folgende: Carry (*fd*), Momentum (*mom*), langfristige Trendumkehr (*q*), Leistungsbilanzsaldo (*ca*), die Produktionslücke (*opg*) sowie der Volatilitätsfaktor (*vol*). Die Optimierung betrachtet eine CRRA-Nutzenfunktion mit  $\gamma = 4$ . Die p-Werte wurden analog zu Barroso und Santa-Clara (2015) über ein Bootstrapping-Verfahren bestimmt. Im Zuge dieses Verfahrens wurden pro Optimierung 1.000 zufällige Datensätze durch Ziehen von Beobachtungen mit Zurücklegen aus dem originalen Datensatz generiert und anschließend für jeden dieser das Optimierungsverfahren erneut durchgeführt. Die Ergebnisse der Koeffizienten wurden auf zwei und jene der p-Werte auf drei Nachkommastellen gerundet. Die Symbole \*, \*\* und \*\*\* kennzeichnen Koeffizienten mit einem p-Wert  $\leq 0,10$ ,  $\leq 0,05$  bzw.  $\leq 0,01$ .

| Strategie     | <i>fd</i>       | <i>mom</i>          | <i>q</i>            | <i>ca</i>           | <i>opg</i>       | <i>vol</i>       |
|---------------|-----------------|---------------------|---------------------|---------------------|------------------|------------------|
| 1. <i>fd</i>  | 14.82***<br>(0) | -                   | -                   | -                   | -                | -                |
| 2. <i>mom</i> | -               | 7.12***<br>(0)      | -                   | -                   | -                | -                |
| 3. <i>q</i>   | -               | -                   | 1.61<br>(0.231)     | -                   | -                | -                |
| 4. <i>ca</i>  | -               | -                   | -                   | 3.42*<br>(0.095)    | -                | -                |
| 5. <i>opg</i> | -               | -                   | -                   | -                   | -2.77<br>(0.959) | -                |
| 6. <i>vol</i> | -               | -                   | -                   | -                   | -                | 1.27<br>(0.214)  |
| 1 - 2         | 18.8***<br>(0)  | 9.46***<br>(0.001)  | -                   | -                   | -                | -                |
| 1 - 3         | 20.02***<br>(0) | 12.63***<br>(0)     | 8.32***<br>(0.002)  | -                   | -                | -                |
| 1 - 4         | 24.56***<br>(0) | 12.48***<br>(0)     | 12.9***<br>(0)      | 12.54***<br>(0.001) | -                | -                |
| 1 - 5         | 24.85***<br>(0) | 12.15***<br>(0)     | 13.06***<br>(0.003) | 12.36***<br>(0.005) | 2.31<br>(0.238)  | -                |
| 1 - 6         | 26.34***<br>(0) | 10.65***<br>(0)     | 12.27***<br>(0.002) | 14.31***<br>(0)     | 3<br>(0.196)     | -5.16<br>(0.954) |
| 1 - 3 & 5     | 20.97***<br>(0) | 12.07***<br>(0.001) | 8.25***<br>(0.003)  | -                   | 4.46*<br>(0.065) | -                |
| 1 - 3 & 6     | 20.22***<br>(0) | 12.4***<br>(0)      | 7.93***<br>(0.02)   | -                   | -                | -1.01<br>(0.664) |

**Tabelle 8:** Analyse der Koeffizienten im Zeitablauf der In-Sample-Periode bei gleitendem Datenfenster

Die Tabelle beruht auf den Charakteristika-Koeffizienten, die zu den Zeitpunkten des In-Sample-Tests (April 1990 bis Oktober 2005) auf Basis der letzten 60 monatlichen Datenpunkte geschätzt und in Abbildung 4 grafisch abgetragen wurden. Namentlich sind es folgende Charakteristika, die sich in den Spaltenüberschriften finden: Carry (*fd*), Momentum (*mom*), langfristige Trendumkehr (*q*), Leistungsbilanzsaldo (*ca*), die Produktionslücke (*opg*) sowie der Volatilitätsfaktor (*vol*). Die ersten beiden Zeilen geben pro Charakteristik Auskunft über den minimalen und maximalen Koeffizienten in der Betrachtungsperiode, während die dritte Zeile basierend darauf die Schwankungsbreite der Koeffizienten darstellt. Zeile vier und fünf zeigen Mittelwert und Standardabweichung der Koeffizienten, bevor in Zeile sechs das Stabilitätsadjustierte Einflussmaß folgt. Letzteres definiert sich in Anlehnung an die Sharpe Ratio als der Quotient aus Mittelwert und Standardabweichung. In der letzten Spalte findet sich der Prozentsatz jener Monate der Betrachtungsperiode, die einen Koeffizienten mit Vorzeichen wider der durch die Literaturrecherche definierten Erwartung aufweisen.

|                                       | <i>fd</i> | <i>mom</i> | <i>q</i> | <i>ca</i> | <i>opg</i> | <i>vol</i> |
|---------------------------------------|-----------|------------|----------|-----------|------------|------------|
| Minimum                               | 26.32     | -8.48      | -6.76    | -6.24     | -24.39     | -41.85     |
| Maximum                               | 89.85     | 22.59      | 46.87    | 92.75     | 56.24      | 3.78       |
| Schwankungsbreite                     | 63.52     | 31.07      | 53.63    | 98.98     | 80.63      | 45.62      |
| Mittelwert                            | 52.45     | 8.68       | 14.28    | 33.15     | 12.64      | -11.54     |
| Standardabweichung                    | 21.25     | 7.12       | 11.69    | 20.65     | 18.99      | 11.79      |
| Stabilitätsadjustiertes Einflussmaß   | 2.47      | 1.22       | 1.22     | 1.61      | 0.67       | 0.98       |
| Prozentsatz inkonsistenter Vorzeichen | 0.00      | 11.02      | 5.51     | 3.94      | 33.07      | 80.31      |



**Abbildung 4:** Veränderung der Koeffizienten im Zeitablauf der In-Sample-Periode bei gleitendem Datenfenster.

Die Abbildung stellt die zu den Zeitpunkten des In-Sample-Tests (April 1990 bis Oktober 2005) auf Basis der letzten 60 monatlichen Datenpunkte geschätzten Koeffizienten der fünften und letzten Kombinationsstrategie dar, welche alle definierten Charakteristika berücksichtigt.

Einfluss auf die Portfoliogewichte (52,45). Danach gliedern sich der Leistungsbilanzsaldo (33,15) und anschließend die langfristige Trendumkehr (14,28) ein. Momentum weist den geringsten absoluten durchschnittlichen Koeffizienten auf. Dass dies nicht voreilig als Zeichen geringer Relevanz von Momentum gedeutet werden sollte, macht das von mir definierte stabilitätsadjustierte Einflussmaß der Charakteristika-Koeffizienten deutlich. Gedanklich angelehnt an die Sharpe Ratio, setzt es die berechneten Mittelwerte der monatlichen Koeffizienten als absolute Werte in Relation zu den entsprechenden Standardabweichungen. Die zugrundeliegende Intuition ist, dass ein Investor zwar jene Charakteristika schätzt, die als Signal gemessen am durchschnittlichen Koeffizienten einen nennenswerten Einfluss auf die Portfoliogewichte im Optimierungsverfahren nehmen, aber zugleich jene scheut, deren Koeffizienten über den Zeitablauf instabil erscheinen, d.h. eine hohe Standardabweichung aufweisen. Carry weist hier den höchsten Wert (2,47) auf, gefolgt vom Leistungsbilanzsaldo (1,61). Den dritten Platz teilen sich Momentum und die langfristige Trendumkehr (1,22). Schlusslichter sind Volatilitätsfaktor (0,98) und Produktionslücke (0,67), die beide auf einen Wert kleiner eins kommen. In Bezug auf den dargelegten Trade-off des Investors ist dies kein gutes Zeichen und bedeutet, dass während der Betrachtungsperiode im Schnitt die monatliche Schwankung der Koeffizienten größer als deren absolute Höhe war.

Zuletzt habe ich in Tabelle 8 pro Charakteristik den Prozentsatz der geschätzten Koeffizienten mit einem Vorzeichen in Widerspruch zur durch die Literaturrecherche definierten Erwartung abgebildet. Die Kennzahl bestätigt in etwa das Bild der vorigen Betrachtungen. Demnach weist kein einziger Carry-Koeffizient ein inkonsistentes Vorzeichen auf, darauf folgend reihen sich Leistungsbilanzsaldo (3,94%), langfris-

tige Trendumkehr (5,51%) und Momentum (11,02%) ein. Produktionslücke (33,07%) und Volatilitätsfaktor (80,31%) bilden wiederum die Schlusslichter.

### 7.2.3. Einordnung und Interpretation der Ergebnisse

Resümierend hat die Einführung des expandierenden Währungsuniversums im Vergleich zum kleineren, statischen Ansatz zu einer wesentlichen Verbesserung der In-Sample-Performance geführt. Dies erscheint logisch gegeben einer Mehrzahl neuer Opportunitäten, aus denen das Verfahren die optimalen Strategien nun zusammensetzen kann.

Weiterhin kann beobachtet werden, dass jene Charakteristika statistisch besser performen, die bereits vergleichsweise lange und in einer gewissen Breite mit wissenschaftlicher Aufmerksamkeit bedacht wurden – mit anderen Worten Carry, Momentum, die langfristige Trendumkehr auf Basis des realen Wechselkurses sowie der Leistungsbilanzsaldo. In diesem Zusammenhang ist zu betonen, dass für diese vier Charakteristika über die betrachtete In-Sample-Periode alle Koeffizienten innerhalb sämtlicher untersuchter Kombinationsstrategien Signifikanz auf dem 1%-Niveau aufwiesen – im In-Sample-Test schafften dies nur Carry und die langfristige Trendumkehr. Zusätzlich zur Signifikanz ist zudem im Schnitt das Ausmaß des Zusammenhangs zwischen Charakteristika und Portfoliogewichten relativ zur In-Sample-Periode gestiegen. Da einzeln betrachtet die Koeffizienten von Leistungsbilanzsaldo, Momentum und Carry den größten Anstieg verzeichnen konnten, liegt der Schluss nahe, dass diese die Dynamik von Währungskursen tendenziell noch besser bei einem breiteren, nicht nur aus Industrienationen bestehendem Anlagespektrum beschreiben.

Dies erscheint sinnvoll, da die Charakteristika zu jedem Zeitpunkt über den Querschnitt aller Währungen standardi-

siert wurden. Hierdurch ist der verwendete Ansatz u.a. wesentlich von einem heterogenen Anlageuniversum im Sinne der selektierten Charakteristika abhängig, da sich sonst fehlende Variabilität bei den Charakteristika in homogenen Portfoliogewichten niederschlägt. Andersgesprochen bietet eine Charakteristik als Signal zur Generierung von Portfolios weniger Mehrwert, wenn sich diese für die betrachteten Währungen kaum unterscheidet. Ein gutes Beispiel hierfür ist die Carry-Charakteristik auf Basis der prozentuellen Forward-Basis, wie methodisch in Abschnitt 3.3.1 definiert. Abbildung 5 zeigt für alle Währungen des Anlagespektrums im In-Sample-Test den durchschnittlichen Wert der Carry-Charakteristik über alle Zeitpunkte der Betrachtungsperiode – sowohl für das statische als auch das expandierende Währungsuniversum.

Zum einen wird beim Betrachten des Abstands zwischen höchstem und niedrigstem Wert direkt eine deutlich höhere Variabilität beim expandierenden Währungsuniversum ersichtlich. Zum anderen wird im Zuge des Rankings klar, dass die hinzugefügten Währungen die vormaligen Long-Währungen, also jene mit einem hohen durchschnittlichen Charakteristik-Wert beim statischen Ansatz, in der Tendenz verdrängt haben. So erreichen von den zehn Währungen mit positivem Vorzeichen des statischen Ansatzes nur noch fünf nach Einführung des expandierenden Universums die gleiche Klassifikation. Einfacher ausgedrückt gestaltet sich die Long-Seite der Carry-Strategie im Schnitt nun dank der hinzugefügten Hochzinsländer interessanter als noch beim statischen Ansatz. Interessant im Zusammenhang mit Carry und dem expandierenden Währungsuniversum ist ebenfalls, dass sowohl Einzelstrategie als auch Benchmark nun eine stark positive Schiefe verzeichnen (2,64 und 3,05). Dieses Ergebnis liefert Evidenz für die Aussage, dass die Überrenditen des Carry-Trades nicht zwangsläufig mit Crash-Risiko einhergehen, wie bspw. [Bekaert und Panayotov \(2018\)](#) in einer aktuellen Arbeit anmerken. Ein bemerkenswerter Aspekt der Ergebnisse ist außerdem, wie sich die Koeffizienten und p-Werte, zunächst ausgehend von den Einzelstrategien, im Rahmen des sukzessiven Hinzufügens neuer Faktoren verhalten. Während die Koeffizienten von Carry und Momentum bereits für sich genommen p-Werte von null ausweisen, erscheint der Koeffizient der langfristigen Trendumkehr separiert betrachtet insignifikant – obwohl dieser beim statischen Universum auch alleine auf einen p-Wert von null kam. In Übereinstimmung damit schwächelt nun auch die ökonomische Performance der besagten Einzelstrategie, während alle anderen Strategien von den zusätzlichen Währungen im Datensatz profitieren. Ein möglicher Grund hierfür könnte in der Inklusion von Entwicklungsländern zum vormalig nur aus Industrienationen bestehenden Universum liegen. Wie in Abschnitt 3.2.2 dargestellt, ist eine wesentliche Annahme in Währungsmodellen auf Basis des realen Wechselkurses dessen Stationarität. [Taylor und Taylor \(2004\)](#) weisen allerdings in diesem Zusammenhang darauf hin, dass der reale Wechselkurs von Entwicklungsländern oft einen Trend aufweist und damit diese Annahme verletzt. Sie schlagen daher eine Trendbereinigung mittels des Harrod-Balassa-Samuelson-

Effekts vor. Da in der vorliegenden Arbeit keine Beseitigung des Trends erfolgte, könnte also das Signal durch Inklusion der Entwicklungsländer an Aussagekraft verloren haben.

Wie in Abschnitt 7.2.2 beschrieben, bewirkt das erstmalige Kombinieren von Carry und Momentum sowie die sukzessive Inklusion von langfristiger Trendumkehr und Leistungsbilanzsaldo zum einen, dass die beiden letztgenannten Charakteristika nun Signifikanz aufweisen, zum anderen aber auch i.d.R. einen Anstieg der Koeffizienten dieser vier Charakteristika, im Unterschied zum statischen In-Sample-Test. Beides spricht für Synergieeffekte bzw. Wechselwirkungen und die steigende Bedeutung einer integrierten Betrachtungsweise der verschiedenen Charakteristika beim expandierenden Währungsuniversum.

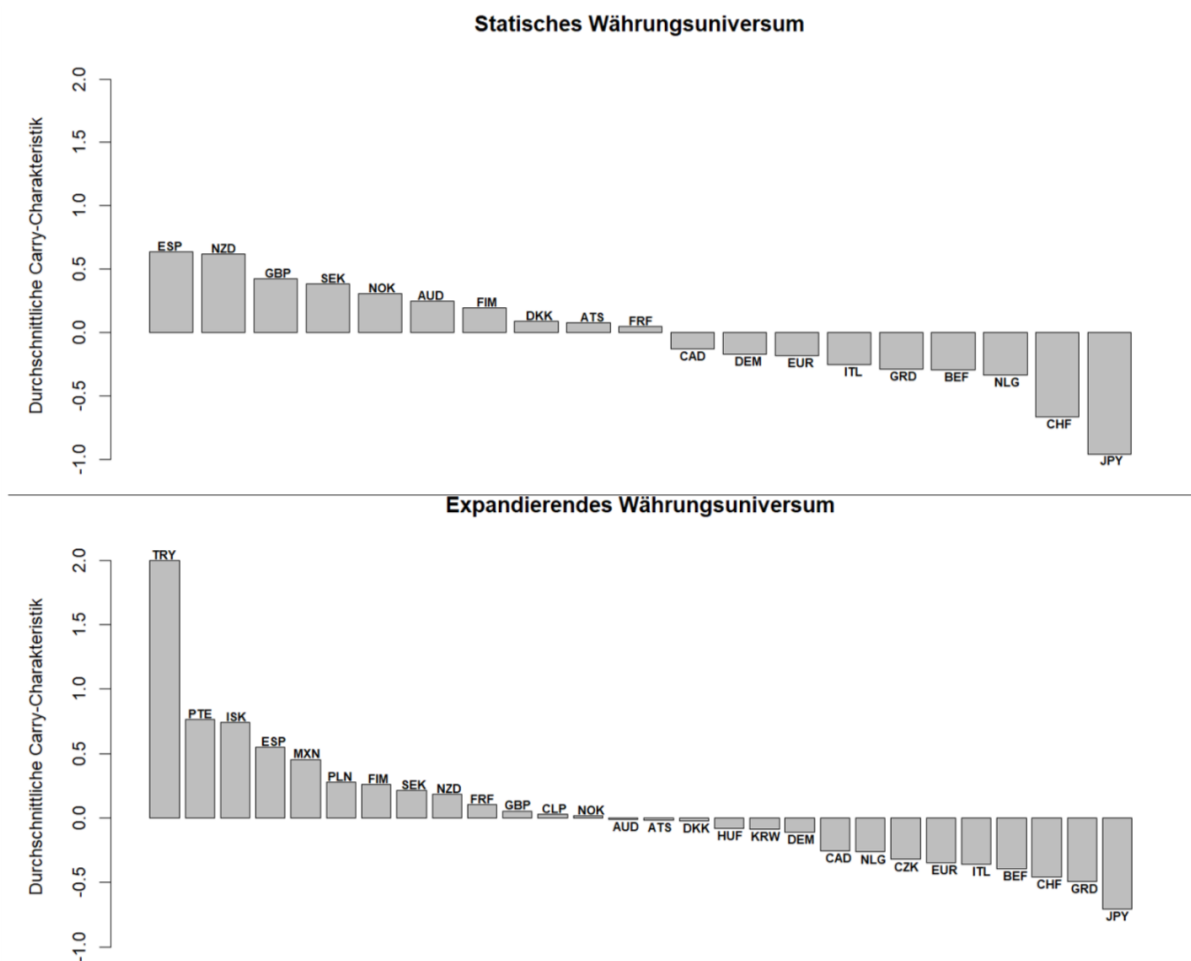
Auf Basis der Analyse mit gleitendem Datenfenster von fünf Jahren können hohe Schwankungsbreiten der Koeffizienten festgestellt werden. Dies lässt darauf schließen, dass der optimale Zusammenhang zwischen den Charakteristika über den Zeitablauf nennenswerten Änderungen unterworfen ist. Im Sinne eines Trade-offs zwischen Einfluss und Stabilität der Koeffizienten können hier ebenfalls Carry, Leistungsbilanzsaldo, Momentum und langfristige Trendumkehr am meisten überzeugen. Zudem verhalten sich diese überwiegend wie in der Literatur dokumentiert.

Interessant gestalten sich zuletzt die Ergebnisse der Korrelationsanalyse, welche nunmehr kein unkorreliertes Gesamtbild aufweist wie beim In-Sample-Test des statischen Universums. Bemerkenswert erscheint hier die hohe Korrelation der beiden Einzelstrategien mit den höchsten Sharpe Ratios – Carry und Momentum – in Höhe von 0,7, welche für die kleinere Länderzahl aus Industrienationen zuvor nur bei 0,08 lag. Daraus erwächst die Vermutung, dass beim expandierenden Universum und einer Querschnittsbetrachtung jene Währungen mit hoher/niedriger prozentueller Forward-Basis auch oft die vergangenen Gewinner/Verlierer der letzten drei Monate waren, beide Einzelstrategien somit im Schnitt über den Zeitverlauf auf einer ähnlichen Allokation basieren und daher die erhöhte Korrelation zustande kommt. Nach Betrachtung der durchschnittlichen Momentum-Charakteristik je Währung wie zuvor in Abbildung 5 für Carry, zu finden in Abbildung F9 des Anhangs F, kann ich dies jedoch nicht bestätigen. Betrachtet man die Schnittmenge der Währungen, die sowohl für Carry- als auch Momentum-Charakteristik durchschnittlich ein positives bzw. negatives Vorzeichen aufweisen, stößt man auf gerade mal fünf bzw. sieben von 28 Ländern. Der massive Anstieg der Korrelation zwischen Carry- und Momentum-Einzelstrategien über den Zeitraum des In-Samples-Test beim expandierenden im Vergleich zum statischen Währungsuniversums bleibt daher eine spannende Frage für weitere Nachforschungen.

### 7.3. Out-of-Sample-Ergebnisse

#### 7.3.1. Performance

Tabelle 9 präsentiert die Performance der optimierten Portfolios sowie jene der definierten Benchmarks über die Out-of-Sample-Zeitperiode von November 2005 bis Dezem-



**Abbildung 5:** Durchschnittliche Werte der Carry-Charakteristik nach Währungen über die In-Sample-Periode.

Die Abbildung zeigt pro verfügbarer Währung im jeweiligen Währungsuniversum als Balken den durchschnittlichen Wert der Carry-Charakteristik, wie sie in Abschnitt 3.3.1 definiert wurde, über den Zeitraum der In-Sample-Periode (April 1990 bis Oktober 2005). Im oberen Teil ist die Analyse auf Basis des statischen, im unteren Teil auf Basis des expandierenden Währungsuniversums abgebildet. Am äußeren Rand der Balken befinden sich zur Identifikation die jeweiligen Währungscode, wie in Abschnitt 4.1 aufgeführt.

ber 2017 bei Verwendung des expandierenden Währungsuniversums – wie zuvor Tabelle 4 in Abschnitt 5.3.1 für das statische Universum. Wie bei den In-Sample-Resultaten hat auch hier die erweiterte Definition des Währungsuniversums im Vergleich zum kleineren statischen Universum merkbare Veränderungen der Ergebnisse bewirkt.

Konnte zuvor im Out-of-Sample-Test nur bei den Einzelstrategien auf Basis von Carry und langfristiger Trendumkehr eine positive Sharpe Ratio beobachtet werden, gesellen sich nun auch Momentum, Leistungsbilanzsaldo und FX-Volatilität hinzu. Die Carry-Einzelstrategie kann dabei mit 0,41 die höchste Sharpe Ratio generieren. Anschließend folgen FX-Volatilität mit 0,19, Leistungsbilanzsaldo mit 0,15, langfristige Trendumkehr mit 0,12 und Momentum mit 0,06. Als letztes ist in diesem Zusammenhang die Produktionslücke mit einer Sharpe Ratio von -0,06 zu nennen. Im Vergleich zu den schwachen In-Sample-Resultaten kann also allgemein eine massive Performance-Verbesserung durch die Einführung des expansiven Währungsuniversums festgestellt werden.

Grund für diese Verbesserung sind teils massiv angestiegene durchschnittliche Renditen bei vergleichsweise moderat angestiegenen Standardabweichungen. Lediglich die Einzelstrategien auf Basis der langfristigen Trendumkehr und Produktionslücke verzeichnen einen Rückgang der mittleren Rendite und nur die langfristige Trendumkehr kann eine verminderte Volatilität aufweisen.

Für das Sicherheitsäquivalent als Performance-Maß ist Ähnliches zu verzeichnen wie bei der Sharpe Ratio. Bei Verwendung des statischen Währungsuniversums waren bisher nur zwei der sechs Einzelstrategien – Volatilitätsfaktor und Leistungsbilanzsaldo – im positiven Bereich, nun schließen sich Carry und langfristige Trendumkehr an. Einfluss hierauf hatte neben den bereits beschriebenen Beobachtungen zusätzlich, dass die Exzess-Kurtosis aller Einzelstrategien mit Ausnahme der langfristigen Trendumkehr gesunken ist und nun alle außer der Produktionslücke eine positive Schiefe aufzeigen.

Im Zuge der Korrelationsanalyse, zu finden in Tabelle F.26 des Anhangs F, kann wie schon zuvor im In-Sample-

**Tabelle 9:** Out-of-Sample-Performance bei expandierendem Währungsuniversum

Die Tabelle stellt die Performance-Kennzahlen der optimierten Strategien sowie der Benchmarks über den Zeitraum der Out-of-Sample-Periode von November 2005 bis Dezember 2017 bei Verwendung des expandierenden Währungsuniversums dar. Die Optimierung betrachtet eine CRRA-Nutzenfunktion mit  $\gamma = 4$ . Aufgeführt werden in den Spalten die maximale ( $\max r$ ), minimale ( $\min r$ ) sowie durchschnittliche ( $\bar{r}$ ) monatliche Rendite, die geschätzte monatliche Standardabweichung ( $\sigma(r)$ ) (alle bisherigen Angaben in Prozent), Exzess-Kurtosis und Schiefe der Wahrscheinlichkeitsverteilung, die monatliche Sharpe Ratio (SR) und das monatliche Sicherheitsäquivalent in Prozent (SE). Die letzten beiden Spalten geben Auskunft über den durchschnittlichen Leverage ( $\bar{L}$ ), angegeben als der Betrag, welcher pro in die risikolose Geldanlage investiertem Dollar in das riskante Währungsportfolio angelegt ist, und Umschlag ( $\bar{T}$ ), der sich als die Summe aus den absoluten Änderungen aller Währungsgewichte des Forward-Portfolios zwischen zwei Perioden definiert. In den ersten sechs Zeilen werden die Kennzahlen für die Einzelstrategien abgebildet. Eine Einzelstrategie bezeichnet dabei das Resultat einer Optimierung, welche nur mit einer einzelnen Charakteristik als Inputfaktor arbeitet. Namentlich sind das folgende Strategien: Carry (fd), Momentum (mom), langfristige Trendumkehr (q), Leistungsbilanzsaldo (ca), die Produktionslücke (opg) sowie der Volatilitätsfaktor (vol). In den nächsten fünf Zeilen folgt die Betrachtung von Strategiekombinationen. Hierfür wurde der Optimierung in einer arbiträren Reihenfolge sukzessiv immer eine weitere Charakteristik als Inputfaktor hinzugefügt, bis die finale Strategie alle vorhandenen Charakteristika miteinbezieht. Wegen der arbiträren Reihenfolge finden sich in den folgenden zwei Zeilen zwei zusätzliche Strategiekombinationen, welche die drei aus der vorangegangenen Analyse relevantesten Charakteristika Carry, Momentum und langfristige Trendumkehr mit jeweils der Produktionslücke und dem Volatilitätsfaktor paaren. In den letzten drei Zeilen der Tabelle findet sich die Performance der drei Benchmarks. Für die Berechnung des Umschlages der US-Aktien Benchmark lagen keine Daten vor.

| Strategie     | $\max r$ | $\min r$ | $\bar{r}$ | $\sigma(r)$ | Kurtosis | Schiefe | SR    | SE    | $\bar{L}$ | $\bar{T}$ |
|---------------|----------|----------|-----------|-------------|----------|---------|-------|-------|-----------|-----------|
| 1. <i>fd</i>  | 71.07    | -26.48   | 7.58      | 18.32       | 2.32     | 1.44    | 0.41  | 2.80  | 9.28      | 8.55      |
| 2. <i>mom</i> | 22.78    | -16.78   | 0.44      | 6.30        | 1.14     | 0.11    | 0.06  | -0.35 | 4.61      | 3.71      |
| 3. <i>q</i>   | 8.38     | -2.35    | 0.26      | 1.40        | 11.38    | 2.75    | 0.12  | 0.22  | 0.95      | 0.25      |
| 4. <i>ca</i>  | 7.80     | -5.91    | 0.42      | 2.18        | 2.10     | 0.83    | 0.15  | 0.33  | 1.91      | 0.18      |
| 5. <i>opg</i> | 7.50     | -10.35   | -0.06     | 2.42        | 2.66     | -0.76   | -0.06 | -0.18 | 1.81      | 0.17      |
| 6. <i>vol</i> | 9.30     | -4.71    | 0.44      | 1.80        | 7.24     | 1.67    | 0.19  | 0.37  | 1.27      | 1.08      |
| 1 - 2         | 89.06    | -35.04   | 9.03      | 21.75       | 1.78     | 1.22    | 0.41  | 2.07  | 12.23     | 11.68     |
| 1 - 3         | 69.58    | -35.8    | 8.14      | 19.77       | 0.97     | 0.80    | 0.41  | 1.57  | 13.89     | 13.06     |
| 1 - 4         | 103.40   | -36.2    | 11.11     | 26.03       | 1.95     | 1.37    | 0.42  | 2.04  | 16.81     | 15.37     |
| 1 - 5         | 98.33    | -36.98   | 10.92     | 25.62       | 1.66     | 1.28    | 0.42  | 1.88  | 16.92     | 15.42     |
| 1 - 6         | 98.05    | -38.86   | 10.74     | 25.81       | 1.57     | 1.23    | 0.41  | 1.31  | 17.07     | 15.59     |
| 1 - 3 & 5     | 63.56    | -37.03   | 7.96      | 19.65       | 0.56     | 0.64    | 0.40  | 1.24  | 14.19     | 13.20     |
| 1 - 3 & 6     | 70.01    | -35.23   | 8.12      | 19.89       | 0.97     | 0.78    | 0.40  | 1.43  | 14.01     | 13.14     |
| Carry         | 99.90    | -35.18   | 10.65     | 25.45       | 2.31     | 1.47    | 0.41  | 2.10  | 12.92     | 11.93     |
| Naiv          | 6.73     | -11.02   | 0.31      | 2.69        | 2.07     | -0.62   | 0.08  | 0.16  | 1.00      | 0.01      |
| US-Aktien     | 11.35    | -17.15   | 0.83      | 4.17        | 2.02     | -0.79   | 0.18  | 0.46  | 0.00      | -         |

Test festgestellt werden, dass die Einzelstrategien mit der besten Performance – hier Carry, Volatilitätsfaktor, Leistungsbilanzsaldo und langfristige Trendumkehr – die vergleichsweise höchsten Korrelationen untereinander aufweisen. Momentum und Produktionslücke kommen bspw. dagegen auf einen negativen Korrelationskoeffizienten mit Carry (-0,04 und -0,55), wirken aber wie beschrieben aufgrund schwacher Performance, etwa ausgedrückt durch Sharpe Ratios von 0,06 und -0,06, relativ uninteressant. Wie bis jetzt in allen Tests kann auch an dieser Stelle Unkorreliertheit mit der US-Aktien-Benchmark bestätigt werden.

Bei den Kombinationsstrategien ist die Performance-Steigerung durch das expandierende Universum ebenfalls zu beobachten. Demnach liegt im Schnitt mehr als eine Vervierfachung der Sharpe Ratios der sukzessiven Kombinationsstrategien relativ zum statischen Ansatz vor, sodass diese nun entweder bei 0,41 oder 0,42 liegen. Als Begründung gilt hier gleiches wie zuvor für die Einzelstrategien, allerdings sind die durchschnittlichen Renditen noch stärker gewachsen, während die Standardabweichungen im Mittel einen ähnlichen Anstieg wie bei den Einzelstrategien verbuchen. Auch die Sicherheitsäquivalente sind nun stark positiv, wobei die Kombinationsstrategien bestehend aus Carry und

Momentum mit 2,07% den höchsten Wert aufweist, dicht gefolgt von jener auf Basis von Carry, Momentum, langfristiger Trendumkehr und Leistungsbilanzsaldo mit 2,04%. Allerdings kann es keine der betrachteten Strategien mit dem Sicherheitsäquivalent der Carry-Einzelstrategien (2,80%) aufnehmen.

Ein vergleichender Blick auf die Benchmarks zeigt ebenfalls ein anderes Bild als in Abschnitt 5.3.1. Konnte noch zuvor keine optimierte Strategie der US-Aktien-Benchmark hinsichtlich Sharpe Ratio und Sicherheitsäquivalent das Wasser reichen, so schlagen nun auf Basis der Sharpe Ratio neun der 13 vorher untersuchten Portfolios und hinsichtlich des Sicherheitsäquivalents zumindest acht die andere Anlageklasse. Die naive Diversifikation wird bzgl. beider Performance-Maße gar von elf der 13 optimierten Strategien übertroffen. Einzig die Carry-Benchmark gibt sich nach wie vor als herausfordernd. In Sachen Sharpe Ratio kann diese es mit fast allen optimierten Strategien aufnehmen; Ausnahmen bilden hier nur die zwei sukzessiven Kombinationsstrategien bis einschließlich Leistungsbilanzsaldo sowie Produktionslücke, welche einen um 0,01 höheren Wert aufweisen. In Bezug auf das Sicherheitsäquivalent kann einzig die Carry-Einzelstrategie die namensgleiche Benchmark übertreffen.



### 7.3.2. Veränderung der Koeffizienten im Zeitablauf

Auf nähere Erläuterungen hinsichtlich der Veränderung der geschätzten Koeffizienten über den Zeitablauf der Out-of-Sample-Periode, welche zuvor in Abbildung 2 des Abschnitts 5.3.2 für das statische Universum präsentiert wurde, wird hier wie zuvor beim In-Sample-Test verzichtet, da sich die Veränderung nicht wesentlich zur vorigen Analyse unterscheidet, auch wenn sich die Koeffizienten i.d.R. nun auf höheren Wertniveaus bewegen. Zur Transparenz findet sich die entsprechende Abbildung F.8 im Anhang E.

Wie im Zuge des In-Sample-Tests habe ich zusätzlich für das expandierende Währungsuniversum untersucht, wie sich die geschätzten Koeffizienten bei monatlicher Schätzung mit einem gleitenden statt des sonst verwendeten expandierenden Datenfensters verändern. Das Ergebnis findet sich in Abbildung 6. Bereits ein Blick auf die Skalierung der y-Achse im Vergleich zur entsprechenden Grafik der In-Sample-Periode verrät, dass die Koeffizienten im Out-of-Sample-Zeitraum innerhalb einer geringeren Bandbreite schwanken.

Des Weiteren habe ich die Analyse aus Tabelle 8 für die Out-of-Sample-Periode repliziert; die Ergebnisse finden sich in Tabelle 10. Die optisch festgestellte Verringerung der Schwankungsbreite im Vergleich zum In-Sample-Test wird durch die quantitative Analyse bestätigt und beträgt im Schnitt ca. 63%.

Bewertet man die Charakteristika nach dem in Abschnitt 7.2.2 definierten stabilitätsadjustierten Einflussmaß des Koeffizienten, belegen nun Leistungsbilanzsaldo und Momentum die letzten beiden Plätze, während diese im Zuge des In-Sample-Tests noch von Volatilität und Produktionslücke eingenommen wurden. Allerdings wird dieses Ergebnis etwa relativiert durch die Tatsache, dass die Vorzeichen der durchschnittlichen Produktionslücke-Koeffizienten überwiegend nicht der Erwartung aus der Literatur entsprechen, im Gegensatz zu den restlichen Charakteristika. Interessanterweise ist im Verhältnis zur In-Sample-Periode der Prozentsatz bei allen Charakteristika angestiegen – nur Carry wirkt mit unveränderten 0% äußerst stabil und einzig der Volatilitätsfaktor kann eine bemerkenswerte Verringerung des Prozentsatzes von 80,31% auf 9,59% verbuchen. Damit konsistent ist nun auch der durchschnittliche Koeffizient jener Charakteristik nicht mehr negativ. Ein Blick in Abbildung 6 zeigt zuletzt, dass der Volatilität-Koeffizient maßgeblich in den letzten zwei Jahren der Out-of-Sample-Periode an Bedeutung gewonnen hat.

### 7.3.3. Einordnung und Interpretation der Ergebnisse

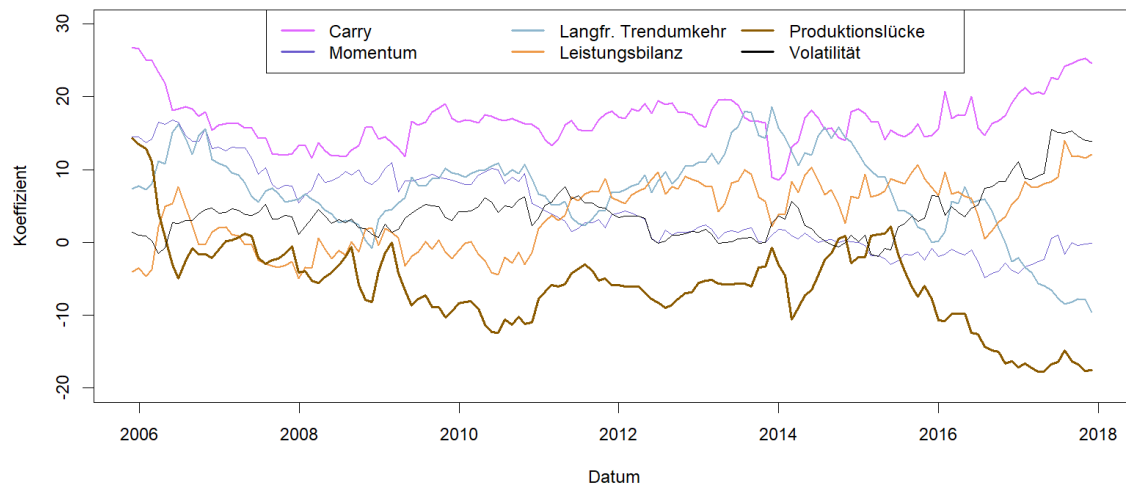
Die wesentliche Performance-Verbesserung im Vergleich zum statischen Währungsuniversum kann auch für die Out-of-Sample-Periode beobachtet werden. Stark angestiegene durchschnittliche Renditen bei lediglich moderat angestiegenen Standardabweichungen im Vergleich zum statischen Universum sind ein Zeichen der neuen Opportunitäten und Diversifikationsvorteile, welche durch die Inklusion der zusätzlichen Währungen auch im Out-of-Sample-Test realisiert werden konnten. Allerdings ist das Bild in Bezug auf die Vorteilhaftigkeit des Ansatzes von Brandt et al. (2009) mit

den zusätzlich zu Carry selektierten Charakteristika auch beim expandierenden Währungsuniversum nicht gänzlich ungetrübt.

Demnach schlagen alle Kombinationsstrategien nun zwar die US-Aktien-Benchmark sowie naive Diversifikation hinsichtlich Sharpe Ratio und Sicherheitsäquivalent. Allerdings kann im Zuge des Out-of-Sample-Tests keine optimierte Strategie das Sicherheitsäquivalent der Carry-Einzelstrategie schlagen und nur zwei schaffen es, deren Sharpe Ratio knapp um 0,01 zu übertreffen, was den Mehrwert des Verfahrens zur Kombination verschiedener Signale bei der Portfolioerstellung und/oder der neben Carry selektierten Charakteristika in Frage stellt. Die Carry-Benchmark bestätigt dieses Bild.

Noch ein anderes interessantes Ergebnis des In- sowie Out-of-Sample-Tests bei Verwendung des expandierenden Währungsuniversums bestärkt die zentrale Rolle von Carry im Kontext von Deviseninvestments. So konnte in beiden Betrachtungsperioden festgestellt werden, dass die Einzelstrategien mit den höchsten Sharpe Ratios nach Carry stets auch die höchsten Korrelationen mit der bekannten Investitionsstrategie aufweisen. In Bezug auf die Korrelation zwischen Carry und der naiven Diversifikation kann in der Out-of-Sample-Periode zudem ein Rückgang von vormals 0,4 beim statischen Ansatz zu nunmehr 0,15 verzeichnet werden. Dies ist in Übereinstimmung mit der vormaligen Interpretation aus Abschnitt 5.3.3, dass die vorherige vergleichsweise hohe Korrelation durch die allgemeine Niedrigzinsphase in den entwickelten Volkswirtschaften, aus denen das statische Universum ausnahmslos bestand, herbei geführt wurde. Die zugrundeliegende Intuition hier war, dass das Signal durch die niedrigen Zinsen an Aussagekraft bei einem relativen Vergleich unter Industrienationen verliert und die Strategie so tendenziell mehr Ähnlichkeit zu einer gleichgewichteten Strategie aufweist. Da nun neue Hochzinswährungen hinzugefügt wurden, ist somit die Korrelation gesunken und das Carry-Portfolio zeigt mehr Variabilität. Des Weiteren weisen wie schon im In-Sample-Test sowohl Carry-Einzelstrategie als auch Carry-Benchmark eine positive Schiefe auf, was gegen die Erklärung der Renditen des Carry-Trades als Kompensation für Crash-Risiko spricht.

Sieht man von der reinen Performance ab und wirft einen Blick auf die Analyse der Koeffizienten im Zuge der Schätzung mit gleitendem Datenfenster von fünf Jahren, überzeugt auch hier Carry als Signal am meisten. So weist sie die zweitniedrigste Schwankungsbreite der Koeffizienten über die Betrachtungsperiode, das mit Abstand höchste stabilitätsadjustierte Einflussmaß und einer Anzahl von null geschätzten Koeffizienten auf, welche einen mit der Literatur inkonsistenten Zusammenhang beschreiben. Die Analyse liefert zudem Evidenz für die Relevanz des realen Wechselkurses bzw. der langfristigen Trendumkehr. So kann die Charakteristik das zweithöchste stabilitätsadjustierte Einflussmaß und den zweitgeringsten Prozentsatz inkonsistenter Vorzeichen realisieren. Interessanterweise kann der Volatilitätsfaktor im Out-of-Sample-Test nicht nur mit der zweithöchsten Performance unter den Einzelstrategien bestehen, sondern weist



**Abbildung 6:** Veränderung der Koeffizienten im Zeitablauf der Out-of-Sample-Periode bei gleitendem Datenfenster.

Die Abbildung stellt die zu den Zeitpunkten des Out-of-Sample-Tests (November 2005 bis Dezember 2017) auf Basis der letzten 60 monatlichen Datenpunkte geschätzten Koeffizienten der fünften und letzten Kombinationsstrategie dar, welche alle definierten Charakteristika berücksichtigt.

**Tabelle 10:** Analyse der Koeffizienten im Zeitablauf der Out-of-Sample-Periode bei gleitendem Datenfenster

Die Tabelle beruht auf den Charakteristika-Koeffizienten, die zu den Zeitpunkten des Out-of-Sample-Tests (November 2005 bis Dezember 2017) auf Basis der letzten 60 monatlichen Datenpunkte geschätzt und in Abbildung 6 grafisch abgetragen wurden. Namentlich sind es folgende Charakteristika, die sich in den Spaltenüberschriften finden: Carry (fd), Momentum (mom), langfristige Trendumkehr (q), Leistungsbilanzsaldo (ca), die Produktionslücke (opg) sowie der Volatilitätsfaktor (vol). Die ersten beiden Zeilen geben pro Charakteristik Auskunft über den minimalen und maximalen Koeffizienten in der Betrachtungsperiode, während die dritte Zeile basierend darauf die Schwankungsbreite der Koeffizienten darstellt. Zeile vier und fünf zeigen Mittelwert und Standardabweichung der Koeffizienten, bevor in Zeile sechs das stabilitätsadjustierte Einflussmaß folgt. Letzteres definiert sich in Anlehnung an die Sharpe Ratio als der Quotient aus Mittelwert und Standardabweichung. In der letzten Spalte findet sich der Prozentsatz jener Monate der Betrachtungsperiode, die einen Koeffizienten mit Vorzeichen wider der durch die Literaturrecherche definierten Erwartung aufweisen.

|                                       | <i>fd</i> | <i>mom</i> | <i>q</i> | <i>ca</i> | <i>opg</i> | <i>vol</i> |
|---------------------------------------|-----------|------------|----------|-----------|------------|------------|
| Minimum                               | 8.59      | -4.85      | -9.60    | -4.94     | -17.85     | -1.91      |
| Maximum                               | 26.85     | 16.80      | 18.62    | 14.03     | 14.28      | 15.47      |
| Schwankungsbreite                     | 18.26     | 21.65      | 28.22    | 18.97     | 32.13      | 17.38      |
| Mittelwert                            | 16.89     | 4.47       | 7.04     | 3.72      | -6.05      | 3.86       |
| Standardabweichung                    | 3.36      | 5.63       | 5.87     | 4.63      | 5.96       | 3.51       |
| Stabilitätsadjustiertes Einflussmaß   | 5.03      | 0.79       | 1.20     | 0.80      | 1.01       | 1.10       |
| Prozentsatz inkonsistenter Vorzeichen | 0.00      | 26.03      | 10.27    | 27.40     | 87.67      | 9.59       |

auch nach Carry und langfristiger Trendumkehr das dritthöchste stabilitätsadjustierte Einflussmaß auf – zudem mit ca. 90% positivem Vorzeichen des Koeffizienten und damit nun im Gegensatz zu den bisherigen Resultaten in Übereinstimmung mit der Erwartung. Demnach scheint es, dass Währungen mit hoher realisierter Volatilität im Verlauf des Vormonats im anschließenden Monat tendenziell eine höhere Rendite aufweisen als jene mit niedriger realisierter Volatilität.

## 8. Transaktionskosten

### 8.1. Änderungen in Methodik und Datensatz

Da die Performance der optimierten Portfolios in der Out-of-Sample-Analyse aus Kapitel 5 nicht an ihren Erfolg im In-Sample-Test anknüpfen konnte, wurde bisher auf die

Betrachtung von Transaktionskosten verzichtet, welche die Performance der Portfolios weiter schmälern sollten. Die Ergebnisse aus Kapitel 7 unter Berücksichtigung des expandierenden Währungsuniversums geben nun Anlass, sich diesem Thema im Weiteren zu widmen.

Sollen im Verfahren zusätzlich Transaktionskosten berücksichtigt werden, kann die Optimierung aus Formel (8) wie folgt erweitert werden:

$$\max_{\theta} \frac{1}{T} \sum_{t=0}^{T-1} u(r_{p,t+1}) = \frac{1}{T} \sum_{t=0}^{T-1} u(r f_{US,t:t+1} + \sum_{i=1}^{N_t} (\frac{\theta^T \hat{x}_{i,t}}{N_t}) r_{i,t+1} - \sum_{i=1}^{N_t} |\frac{\theta^T \hat{x}_{i,t}}{N_t}| c_{i,t+1}), \quad (22)$$

$c_{i,t+1}$  stellt dabei die Transaktionskosten von Währung  $i$  zum Zeitpunkt  $t$  dar und bestimmt sich folgendermaßen:

$$c_{i,t+1} = \frac{F_{i,t:t+1}^{ask} - F_{i,t:t+1}^{bid}}{F_{i,t:t+1}^{ask} + F_{i,t:t+1}^{bid}}. \quad (23)$$

Demnach sind die Transaktionskosten analog zu Barroso und Santa-Clara (2015) als die Hälfte des Ask-Bid-Spreads in Prozent vom Mid-Preis der jeweiligen Währung definiert. Dies beruht auf der Annahme, dass der hypothetische Investor zum Ask-Preis (Bid-Preis) Währungen im Forward-Markt kauft (verkauft) und der Kontrakt in der anschließenden Periode zum Spot-Kurs abgerechnet wird. Barroso und Santa-Clara (2015) deuten darauf hin, dass dieses Vorgehen die Transaktionskosten wohl tendenziell eher überschätzt. Sie verweisen im Zuge dessen auf Mancini et al. (2013), welche im Spot-Markt die effektiven Kosten geringer als die Hälfte der durch Bid-Ask-Preise implizierten Kosten dotieren.

Zur Erlangung von Ask- und Bid-Forward-Kursen habe ich Bloomberg genutzt und konnte so für alle in Abschnitt 7.1 aufgeführten 28 Länder des expandierenden Datensatzes Zeitreihen beziehen.<sup>32</sup> Da diese Reihen allerdings frühestens im Jahr 1996 – sechs Jahre nach Beginn der Gesamtbeobachtungsperiode – starten, wurden sie mittels einer Durchschnittsbetrachtung zurückgerechnet. Hierfür habe ich zunächst für alle Währungen zu jedem Zeitpunkt die prozentuelle Abweichung des Ask- bzw. Bid-Kurses vom Mid-Kurs berechnet, falls die entsprechenden Daten zur Verfügung standen. Aus diesen Angaben habe ich für sämtliche Währungen die durchschnittliche prozentuelle Abweichung berechnet und diese auf die Mid-Kurse aller Datenpunkte mit fehlenden Werten angewandt, um so die Ask- und Bid-Zeitreihen aufzufüllen.<sup>33</sup>

Bei der Performance-Analyse betrachte ich in einem Zwischenschritt zunächst, wie die optimalen Portfolios aus Kapitel 7 nach Berücksichtigung von Transaktionskosten performen, bevor ich im nächsten Schritt wie vorgestellt die Kosten auch im Optimierungsverfahren inkludiere.

## 8.2. Ergebnisse und Interpretation

Die Ergebnisse der durchgeführten Analyse befinden sich in Anhang G. Tabelle G.27 quantifiziert die In-Sample-Performance und zeigt im ersten Schritt, dass die Sharpe Ratios der optimierten Strategien aus Abschnitt 7.2.1 durch die Berücksichtigung von Transaktionskosten in unterschiedlichem Ausmaß geschmälert werden. Während die

entsprechenden Kennzahlen der Einzelstrategien auf Basis von langfristiger Trendumkehr, Leistungsbilanzsaldo und FX-Volatilität – vormals alle mit positiven Sharpe Ratios – einen Vorzeichenwechsel verzeichnen und jene der Produktionslücke nun in etwa bei null liegt, wirken die Sharpe Ratios von Momentum und Carry (0,41 und 0,21) auch nach Berücksichtigung von Transaktionskosten im Vergleich zu den Benchmarks noch attraktiv. Grund hierfür ist, dass bei den beiden letztgenannten Strategien die Mittelwerte bzw. Standardabweichungen durch die Berücksichtigung von Transaktionskosten prozentual am geringsten sinken bzw. steigen. Zu bemerken ist auch, dass die Carry-Einzelstrategie die Einführung der Kosten wesentlich besser verträgt als die gleichnamige Benchmark, obwohl beide zuvor sowohl in Hinblick auf Sharpe Ratio als auch Sicherheitsäquivalent ähnlich positioniert waren. Die durchschnittliche Rendite und Sharpe Ratio der Benchmark nach Kosten sind demnach nun negativ. Ein Vorteil der integrierten Betrachtung verschiedener Inputsignale als Strategiekombinationen wird, wie bereits auch Barroso und Santa-Clara (2015) feststellen, bei Berücksichtigung von Transaktionskosten deutlich. So ist der stärkste prozentuale Rückgang unter den Sharpe Ratios der Kombinationsstrategien (10,64% bei der Kombination aus Carry, Momentum und langfristiger Trendumkehr) geringer als der kleinste prozentuale Rückgang unter denen der Einzelstrategien (12,77% bei Carry). Auch das Sicherheitsäquivalent geht im Durchschnitt bei den sukzessiven Kombinationsstrategien nur um knapp 35% zurück, während es bei den Einzelstrategien um etwa 57% sinkt.

Tabelle G.27 gibt zudem Aufschluss darüber, wie sich die Performance der Strategien innerhalb der In-Sample-Periode gestaltet, wenn die Transaktionskosten wie im vorherigen Abschnitt beschrieben in das Optimierungsverfahren einfließen. Alle Strategien können wenig überraschend Steigerungen in den Performance-Maßen im Vergleich zu den Ergebnissen des vorherigen Absatzes verbuchen. Zudem lässt sich hier ein gegenläufiger Trend erkennen: Die Einzelstrategien, welche im Zuge der Einführung von Transaktionskosten bei der Performance-Ermittlung die größten Einbußen aufwiesen, zeigen nun die stärkste Performance-Regeneration. Bei den Kombinationsstrategien hingegen, die weniger stark unter der Einführung der Transaktionskosten gelitten hatten, fällt der Performance-Aufschwung moderater aus. Betrachtet man in Tabelle G.28, wie sich die geschätzten Koeffizienten durch die Inklusion des Terms der Transaktionskosten geändert haben, können keine wesentlichen Unterschiede, aber überwiegend eine Annäherung zu null beobachtet werden. Dementsprechend sind auch die p-Werte in der Mehrzahl leicht gestiegen, aber alle Koeffizienten, welche zuvor bereits auf dem 5%-Niveau signifikant waren, halten diesen Status weiterhin.

In Tabelle G.29 präsentiere ich schließlich die Performance-Kennzahlen über die Out-of-Sample-Periode. Es zeigt sich, dass die In-Sample-Resultate halten. Demnach nehmen Transaktionskosten auf die in Abschnitt 7.3.1 dargestellten Performance-Kennzahlen einen wesentlichen negativen Einfluss – der Rückgang ist wiederum stärker für die Einzel-

<sup>32</sup>Hierbei wurden die gleichen Bloomberg-Ticker wie für die Mid-Zeitreihen verwendet, zu finden in Tabelle A.13 des Anhangs A – lediglich das Bloomberg-Feld wurde von PX\_LAST auf PX\_ASK bzw. PX\_BID umgestellt.

<sup>33</sup>Das Vorgehen impliziert, dass es sich bei den auf Basis der nachträglich aufgefüllten Ask- und Bid-Werte errechneten Transaktionskosten  $c_{i,t+1}$  um eine Konstante je Währung handelt, nämlich der durchschnittlichen Hälfte des Ask-Bid-Spreads in Prozent vom Mid-Preis der jeweiligen Währung. Um die Modellierung möglichst einfach und gut nachvollziehbar zu halten, wird bspw. von einem etwaigen Trend in der Entwicklung der Transaktionskosten über die Zeit abstrahiert.

als die Kombinationsstrategien zu beobachten, wobei etwa die Sharpe Ratio von Momentum nun negativ ist, jene der langfristigen Trendumkehr sich halbiert und sich Carry wiederum am geringsten betroffen zeigt. Bei der folgenden Inklusion der Transaktionskosten in das Optimierungsverfahren kann auch für die Out-of-Sample-Periode allgemein eine Performance-Regeneration konstatiert werden – wie zuvor wiederum stärker für die Einzel- als die Kombinationsstrategien.

Zusammenfassend stelle ich wie schon Barroso und Santa-Clara (2015) zuvor fest, dass Transaktionskosten in Form von Bid- und Ask-Spreads eine wesentliche Belastung für die Performance von Deviseninvestments im Forward-Markt sein können. Nichtsdestotrotz und ebenfalls wie bei Barroso und Santa-Clara (2015) bleibt die Outperformance der optimierten Portfolios auch nach Berücksichtigung der Transaktionskosten bestehen, wenn diese auch Eingang in das Optimierungsverfahren finden. Da der durchschnittliche Leverage der Strategien im Zuge der Inklusion von Transaktionskosten zwar bereits etwas sinkt, allerdings vor allem für die Kombinationsstrategien im Schnitt mit einem zweistelligen Hebel weiterhin sehr hoch ist, wäre eine nächste spannende Erweiterung des Ansatzes für zukünftige Arbeiten, Leverage innerhalb der Optimierung zu limitieren und/oder mit zusätzlichen Kosten zu bestrafen.

## 9. Conclusio

Im einleitenden Kapitel habe ich einen kurzen Einblick in die historische Entwicklung der FX-Forschungsliteratur gegeben, die in starkem Zusammenhang mit der Frage nach der Effizienz von Devisenmärkten, dem empirischen Versagen der ungedeckten Zinsparität und der daraus abgeleiteten Anlagestrategie des Carry-Trades steht. Danach beschäftigte sich das zweite Kapitel mit der Frage nach der optimalen Portfolioallokation und führte die Methodik des Verfahrens von Brandt et al. (2009) zur Bestimmung der optimalen Portfoliogewichte in Anlehnung an Barroso und Santa-Clara (2015) im Devisenkontext ein. Zudem habe ich hier die Performance-Benchmarks definiert, mit denen sich die optimierten Währungsstrategien im Weiteren messen mussten, namentlich der diversifizierte Carry-Trade, die naive Diversifikation und der US-Aktienmarkt.

Da der Ansatz von Brandt et al. (2009) zur Modellierung der Dynamik der optimalen Portfoliogewichte auf sogenannte Charakteristika als Signale zurückgreift, statt die Allokation wie bspw. das bekannte Verfahren von Markowitz (1952) auf der Schätzung erwarteter Renditen und Standardabweichungen zu fundieren, selektierte ich im dritten Kapitel zur weiteren Verwendung im empirischen Teil zusätzlich zu Carry eine Auswahl von fünf Charakteristika auf Basis einschlägiger Forschungsliteratur. Im Folgenden flossen demnach auch Momentum, der reale Wechselkurs, der Leistungsbilanzsaldo sowie Produktionslücke und FX-Volatilität als Signale in das Optimierungsverfahren ein.

Nachdem ich in Kapitel 4 Aufschluss über die zugrundeliegende Datenbasis gegeben habe, präsentierte ich im fünften

ten Kapitel zunächst die In- und Out-of-Sample-Ergebnisse bei Betrachtung eines kleinen statischen Universums, lediglich bestehend aus Industrienationen mit mindestens 25 Jahren an Daten pro benötigtem Datentyp. Anschließend untersuchte ich in Kapitel 6 im Zuge einer Szenarioanalyse, wie sich die Resultate des vorherigen Kapitels bei Variation der angenommenen relativen Risikoaversion innerhalb des Ansatzes verändern. Im nächsten Schritt habe ich im siebten Kapitel die Definition des Währungsuniversums erweitert und betrachtete fortan alle Währungen, für die sämtliche nötigen Daten zur Berechnung der Charakteristika vorlagen. Demnach wurde eine Währung initial in das Universum inkludiert, sobald für alle Datentypen Werte vorlagen. Zuletzt betrachtete Kapitel 8 auf Basis der Ergebnisse des vorherigen Kapitels, wie robust die Ergebnisse sich bei zusätzlicher Betrachtung von Transaktionskosten zeigten. Die dabei zustande gekommenen wesentlichen empirischen Resultate lauten wie folgt:

- Carry wird der eingangs beschriebenen Aufmerksamkeit durch die wissenschaftliche Literatur gerecht. Demnach erweisen sich sämtliche im Zuge der diversen Signifikanzanalysen betrachteten Koeffizienten für die besagte Charakteristik als signifikant auf dem 1%-Niveau und auch die ökonomische Out-of-Sample-Performance, gemessen an Sharpe Ratio und Sicherheitsäquivalent, überzeugt am meisten. Auch erweist sich der Zusammenhang zwischen Carry-Signal und optimalen Portfoliogewichten bei Analyse mit gleitendem Datenfenster von allen untersuchten Charakteristika am stabilsten. Zudem kann ich bei Betrachtung des expandierenden Währungsuniversums für die bekannte Strategie eine positive Schiefe verzeichnen, was gegen eine allgemeingültige Erklärung der Überrenditen als Kompensation für Crash-Risiko spricht.
- Ebenfalls überwiegend signifikant auf dem 5%-Niveau zeigen sich die Koeffizienten des realen Wechselkurses, Leistungsbilanzsaldos und von Momentum; dies steht in Einklang mit der weitreichenden Dokumentation dieser Faktoren in der Literatur. Die Produktionslücke kann in keiner der verschiedenen Analysen überzeugen und obschon der Volatilitätsfaktor mitunter eine interessante Performance verbuchen kann, scheint der Zusammenhang instabil.
- Im Zuge der Analyse verschiedener Szenarien mit unterschiedlicher Risikoaversion wird deutlich, dass die Koeffizienten der genannten Charakteristika ihre Signifikanz bei steigender Risikoaversion zwar behalten, der hypothetische Investor mit der CRRA-Nutzenfunktion aber ein deutlich geringeres Exposure bevorzugt. Dies lässt darauf schließen, dass die durch signifikante Charakteristika generierten Gewinne vom Investor mit Risiko assoziiert werden und keinesfalls einfach für anomal höhere durchschnittliche Renditen stehen. Demnach sind vornehmlich Investoren mit Risikoappetit und -tragfähigkeitspotential bei der Realisierung



besagter Gewinne gefragt.

- Die Ergebnisse machen deutlich, dass die Profitabilität beinahe aller untersuchten Strategien eine maßgebliche Verbesserung durch eine Erweiterung der Anlageopportunitäten in Form neuer Währungen erfährt, sobald die notwendigen Daten für deren Inklusion in das Optimierungsverfahren bereitstehen. Vor allem über die stark durch die Niedrigzinsphase in den entwickelten Volkswirtschaften geprägte Out-of-Sample-Periode von November 2005 bis Dezember 2017 verliert bspw. Carry an Relevanz, wenn Entwicklungsländer mit höherem Zinsniveau ausgeschlossen werden und das Signal dadurch Variabilität einbüßt. Lediglich der reale Wechselkurs schneidet für sich allein genommen bei der Einführung des expandierenden Währungsuniversums schlechter ab als zuvor, was mit der Inklusion von Entwicklungsländern in Zusammenhang stehen mag, deren realer Wechselkurs oft gegen die angenommene Stationarität ökonomischer Modelle verstößt. Eine Trendbereinigung könnte in diesem Fall die Aussagekraft des Faktors verbessern. Trotz allem kann der reale Wechselkurs in Kombination mit den anderen Charakteristika auch ohne etwaige Trendbereinigung beim expandierenden Währungsuniversum weiterhin überzeugen.
- Es zeigt sich, dass die untersuchten Charakteristika teilweise interessante Diversifikationseigenschaften und positive Wechselwirkungen aufweisen, welche sich durch Kombination mittels des Optimierungsansatzes von [Brandt et al. \(2009\)](#) realisieren lassen. Besonders im Zuge der Einführung von Transaktionskosten erweist sich demnach die Performance der Kombinationsim Vergleich zu den Einzelstrategien als robuster.
- Transaktionskosten haben nennenswerte negative Auswirkungen auf die Performance der untersuchten Strategien. Eine Integration als Variable in das Optimierungsverfahren führt allerdings zu einer Regeneration der Performance.
- Das unkorrelierte Gesamtbild, welches die FX-Einzelstrategien in allen Analysen mit der US-Aktien-Benchmark aufweisen, deutet auf vielversprechende Diversifikationseigenschaften für einen (US-)Aktieninvestor durch die zusätzliche Betrachtung der Assetklasse hin.

Auch wenn damit schon einige Erkenntnisse gewonnen werden konnten, bleibt wie immer Raum für weitere Analysen. Demnach könnte in einem nächsten Schritt der bisher sehr hohe Leverage der optimierten Strategien limitiert und/oder mit zusätzlichen Kosten bestraft werden. Da der wesentliche Einfluss eines breiten Anlagespektrums auf die Modell-Performance beobachtet werden konnte, gestaltet sich auch die Verbreiterung der Datenbasis und eine erneute Erweiterung des Währungsuniversums interessant. Zuletzt liefert die wissenschaftliche Literatur noch weitere interessante Charakteristika, die in die präsentierte Analyse

integriert werden könnten, aber auch alternative Definitionen und Ansätze zu den bereits untersuchten Charakteristika für einen interessanten Vergleich. Als finale Erkenntnis kann demnach festgestellt werden, dass die Forschungsmöglichkeiten im Kontext von Devisenanlagestrategien weiterhin versprechen interessant zu bleiben.



## Literatur

- Adrian, T. und Rosenberg, J. Stock returns and volatility: Pricing the short-run and long-run components of market risk. *The Journal of Finance*, 63 (6):2997–3030, 2008.
- Akram, Q. F., Rime, D., und Sarno, L. Arbitrage in the foreign exchange market: Turning on the microscope. *Journal of International Economics*, 76(2):237–253, 2008.
- Ang, A., Hodrick, R. J., Xing, Y., und Zhang, X. The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299, 2006.
- Asness, C. S., Moskowitz, T. J., und Pedersen, L. H. Value and momentum everywhere. *The Journal of Finance*, 68(3):929–985, 2013.
- Balduzzi, P. und Chiang, I.-H. E. Real Exchange Rates and Currency Risk Premia, 2017.
- Bank für Internationalen Zahlungsausgleich. Triennial Central Bank Survey - Foreign exchange turnover in April 2016, 2016.
- Barberis, N., Shleifer, A., und Vishny, R. A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343, 1998.
- Barroso, P. und Santa-Clara, P. Beyond the Carry Trade: Optimal Currency Portfolios, 2015. *Journal of Financial and Quantitative Analysis* 50.
- Bekaert, G. und Panayotov, G. Good Carry, Bad Carry, 2018.
- Bhansali, V. Volatility and the carry trade. *The Journal of Fixed Income*, 17 (3):72–84, 2007.
- Black, F. und Litterman, R. Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43, 1992.
- Bliss, R. R. und Panigirtzoglou, N. Option-implied risk aversion estimates. *The Journal of Finance*, 59(1):407–446, 2004.
- Brandt, M. W. Portfolio Choice Problems, in Yacine Ait-Sahalia, 2010.
- Brandt, M. W., Santa-Clara, P., und Valkanov, R. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, 22(9):3411–3447, 2009.
- Britten-Jones, M. und Neuberger, A. Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, 55(2):839–866, 2000.
- Brunnermeier, M. K., Nagel, S., und Pedersen, L. H. Carry trades and currency crashes. *NBER Macroeconomics Annual*, 23(1):313–348, 2008.
- Burda, M. C. und Wyplosz, C. Makroökonomie. Eine europäische Perspektive, 2009.
- Burnside, C., Rebelo, S., und Eichenbaum, M. Carry trade: The gains of diversification, Proceedings of the Twenty-Second Annual Congress of the European Economic Association, 2008.
- Burnside, C., Eichenbaum, M., und Rebelo, S. Carry trade and momentum in currency markets. *Annu. Rev. Financ. Econ.*, 3(1):511–535, 2011.
- Burstein, A. und Gopinath, G. International Prices and Exchange Rates, in Gita Gopinath, Elhanan Helpman, und Kenneth S. Rogoff, eds, 2014.
- Cheung, Y.-W., Chinn, M. D., und Pascual, A. G. Empirical exchange rate models of the nineties: Are any fit to survive? *Journal of International Money and Finance*, 24(7):1150–1175, 2005.
- Clarida, R., Davis, J., und Pedersen, N. Currency carry trade regimes: Beyond the fama regression. *Journal of International Money and Finance*, 28(8):1375–1389, 2009.
- Cochrane, J. H. The dog that did not bark: A defense of return predictability. *The Review of Financial Studies*, 21(4):1533–1575, 2007.
- Cochrane, J. H. Presidential address: Discount rates. *The Journal of Finance*, 66(4):1047–1108, 2011.
- Cochrane, J. H. Macro-finance. *Review of Finance*, 21(3):945–985, 2017.
- Corte, P. D., Riddiough, S. J., und Sarno, L. Currency premia and global imbalances. *The Review of Financial Studies*, 29(8):2161–2193, 2016.
- Croushore, D. und Stark, T. A real-time data set for macroeconomists: Does the data vintage matter? *Review of Economics and Statistics*, 85(3):605–617, 2003.
- Dahlquist, M. und Penasse, J. The Missing Risk Premium in Exchange Rates, 2017.
- Daniel, K., Hirshleifer, D., und Subrahmanyam, A. Investor psychology and security market under- and overreactions. *Journal of Finance*, 53(6):1839–1885, 1998.
- Della Corte, P., Ramadorai, T., und Sarno, L. Volatility risk premia and exchange rate predictability. *Journal of Financial Economics*, 120(1):21–40, 2016.
- DeMiguel, V., Garlappi, L., und Uppal, R. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953, 2007.
- Dornbusch, R. und Fischer, S. Exchange rates and the current account. *The American Economic Review*, 70(5):960–971, 1980.
- Duchin, R. und Levy, H. Markowitz versus the talmudic portfolio diversification strategies. *Journal of Portfolio Management*, 35(2):71, 2009.
- Engel, C. und West, K. D. Global interest rates, currency returns, and the real value of the dollar. *American Economic Review*, 100(2):562–67, 2010.
- Fama, E. F. Forward and spot exchange rates. *Journal of Monetary Economics*, 14(3):319–338, 1984.
- Fama, E. F. und French, K. R. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.
- Filipe, S. F. und Suominen, M. Currency Carry Trades and Funding Risk, AFA 2014 Philadelphia Meetings, American Finance Association, 2014.
- Fisher, I. Publications of the American Economic Association, 1896. 1–98.
- Frankel, J. A. und Rose, A. K. Empirical research on nominal exchange rates, in Gene M. Grossman, und Kenneth S. Rogoff, eds, 1995.
- Friedman, M. *Essays in positive economics*. University of Chicago Press, 1953.
- Froot, K. A. und Ramadorai, T. Currency returns, intrinsic value, and institutional-investor flows. *The Journal of Finance*, 60(3):1535–1566, 2005.
- Froot, K. A. und Thaler, R. H. Anomalies: foreign exchange. *Journal of Economic Perspectives*, 4(3):179–192, 1990.
- Fung, W. und Hsieh, D. A. Hedge fund benchmarks: A risk-based approach. *Financial Analysts Journal*, 60(5):65–80, 2004.
- Gabaix, X. und Maggiori, M. International liquidity and exchange rate dynamics. *The Quarterly Journal of Economics*, 130(3):1369–1420, 2015.
- Giordani, P. und Halling, M. Valuation ratios and shape predictability in the distribution of stock returns. *Swedish House of Finance Research Paper*, (17-5), 2019.
- Gourinchas, P.-O. und Rey, H. International financial adjustment. *Journal of Political Economy*, 115(4):665–703, 2007.
- Hansen, L. P. und Hodrick, R. J. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy*, 88(5):829–853, 1980.
- Hong, H. und Stein, J. C. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 54 (6):2143–2184, 1999.
- Hooper, P. und Morton, J. Fluctuations in the dollar: A model of nominal and real exchange rate determination. *Journal of International Money and Finance*, 1:39–56, 1982.
- Jacobs, H., Müller, S., und Weber, M. How should individual investors diversify? an empirical evaluation of alternative asset allocation policies. *Journal of Financial Markets*, 19:62–85, 2014.
- Jegadeesh, N. und Titman, S. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91, 1993.
- Jordà, Ò. und Taylor, A. M. The carry trade and fundamentals: Nothing to fear but fear itself. *Journal of International Economics*, 88(1):74–90, 2012.
- Jurek und Jakub W., . Crash-neutral currency carry trades. *AFA 2010 Atlanta Meetings Paper*, American Finance Association, 2010.
- Jylhä, P. und Suominen, M. Speculative capital and currency carry trades. *Journal of Financial Economics*, 99(1):60–75, 2011.
- Korajczyk, R. A. The pricing of forward contracts for foreign exchange. *Journal of Political Economy*, 93(2):346–368, 1985.
- Kouri, P. J. The exchange rate and the balance of payments in the short run and in the long run: A monetary approach. *The Scandinavian Journal of Economics*, pages 280–304, 1976.
- Kroencke, T. A., Schindler, F., und Schrimpf, A. International diversification benefits with foreign exchange investment styles. *Review of Finance*, 18 (5):1847–1883, 2013.
- Lee, J. und Chinn, M. D. Current account and real exchange rate dynamics in the G7 countries. *Journal of International Money and Finance*, 25(2):257–274, 2006.
- Lewis, K. K. Changing beliefs and systematic rational forecast errors with evidence from foreign exchange. *The American Economic Review*, pages 621–636, 1989.
- Lintner, J. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets, 1965.
- Lothian, J. R. und Taylor, M. P. Real exchange rate behavior: the recent float from the perspective of the past two centuries. *Journal of Political Economy*, 104(3):488–509, 1996.

- Lustig, H., Roussanov, N., und Verdelhan, A. Common risk factors in currency markets. *The Review of Financial Studies*, 24(11):3731–3777, 2011.
- Lustig, H., Stathopoulos, A., und Verdelhan, A. The Term Structure of Currency Carry Trade Risk Premia (National Bureau of Economic Research, Cambridge, MA), 2017.
- Mancini, L., Rinaldo, A., und Wrampelmeyer, J. Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums. *The Journal of Finance*, 68(5):1805–1841, 2013.
- Markowitz, H. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.
- Meeße, R. A. und Rogoff, K. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14(1-2): 3–24, 1983.
- Menkhoff, L. und Taylor, M. P. The obstinate passion of foreign exchange professionals: technical analysis. *Journal of Economic Literature*, 45(4): 936–972, 2007.
- Menkhoff, L., Sarno, L., Schmeling, M., und Schrimpf, A. Carry trades and global foreign exchange volatility. *The Journal of Finance*, 67(2):681–718, 2012a.
- Menkhoff, L., Sarno, L., Schmeling, M., und Schrimpf, A. Currency momentum strategies. *Journal of Financial Economics*, 106(3):660–684, 2012b.
- Mussa, M. A model of exchange rate dynamics. *Journal of Political Economy*, 90(1):74–104, 1982.
- Nucera, F. Unemployment fluctuations and the predictability of currency returns. *Journal of Banking & Finance*, 84:88–106, 2017.
- Nurkse, R. International Currency Experience. Lessons of the Inter-War Period, Völkerbund, 1944.
- Obstfeld, M. Financial Risk: Theory, Evidence and Implications. Proceedings of the Eleventh Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis, 1989.
- Okunev, J. und White, D. Do momentum-based strategies still work in foreign currency markets? *Journal of Financial and Quantitative Analysis*, 38(2):425–447, 2003.
- Riddiough, S. J. und Sarno, L. Business Cycles and the Cross-Section of Currency Returns, 2017.
- Rogoff, K. The purchasing power parity puzzle. *Journal of Economic literature*, 34(2):647–668, 1996.
- Shiller, R. J. Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?, 1981. 421–436.
- Taylor, A. M. und Taylor, M. P. The purchasing power parity debate. *Journal of Economic Perspectives*, 18(4):135–158, 2004.
- Taylor, D. Official intervention in the foreign exchange market, or, bet against the central bank. *Journal of Political Economy*, 90(2):356–368, 1982.
- Tobin, J. A general equilibrium approach to monetary theory. *Journal of Money, Credit and Banking*, 1(1):15–29, 1969.
- Van Binsbergen, J. H. und Koijen, R. S. Predictive regressions: A present-value approach. *The Journal of Finance*, 65(4):1439–1471, 2010.
- Verdelhan, A. A habit-based explanation of the exchange rate risk premium. *The Journal of Finance*, 65(1):123–146, 2010.