Kann mit Socially Responsible Investing (SRI) eine Überrendite erzielt werden?

Bachelor Thesis in Banking and Finance

Pascal Patrik Boll

Full Text Version

CMF Thesis Series no. 21 (2016)

July 2016

Center for Microfinance Thesis Series

Title Kann mit Socially Responsible Investing (SRI) eine Überrendite erzielt werden?
Bachelor Thesis in Banking and Finance

Pascal Patrik Boll
Advisor: Dr. Annette Krauss
Professor: Professor Dr. Marc Chesney

Full Text Version

Executive Summary

Problem

Environmental, social, and governance issues have gained significant importance in recent years. Global warming and natural catastrophes such as the leaking oil platform Deep Water Horizon are examples of the first issue. Besides these, working conditions, employee satisfaction, donations and other social matters have become central points for companies' steering committees. Further, excess salaries and bonuses have raised anger within society, which has made remuneration and corporate governance policies publicly discussed topics; the Minder initiative in Switzerland illustrates this development well. It is clear that the financial sector, as a central economic player and responsible for efficient resource allocation within the economy, cannot escape this dynamic. The awareness of society of these problems is also noticeable among investors, including institutional investors. Investment volume in the investment style referred to as socially responsible investing (SRI) has increased at double-digit rates in the last decade. Between 2012 and 2014 alone, sustainable managed assets increased by 61%. This remarkable development raises the question: Is it possible to do good while doing well or even doing better? This study investigates whether it is possible to generate a superior return by implementing a SRI strategy or, viewed from a broader perspective, whether sustainable companies achieve a better financial performance.

Procedure

To tackle these questions, the study takes three steps. The first chapter presents a general overview of SRI, important definitions, and the development of SRI. Moreover, a theoretical approach to this topic is presented and the six major existing theoretical frameworks that address the relationship between corporate social and financial performance are explained. In the second chapter, the current state of secondary literature about the linkage between corporate social performance (CSP) and corporate financial performance (CFP) is evaluated qualitatively with implications for the reliability of studies and their results. In accordance with the findings of the preliminary literature review, an empirical stock-market study is presented in the third chapter. In this study, two differently weighted best-in-class portfolios are constructed employing newly available environmental, social, and governance (ESG) data from Sustainalytics. The risk-adjusted performance (controlled with Fama-French factors) is observed over a three-year period in hindsight: between April 2013 and April 2016.

Results and Overall Evaluation

The review of the secondary papers reveals four types of secondary studies: narrative, theoretical, vote-count, and meta-analysis studies. The studies and their results differ in terms of their expressiveness, which is basically due to the diversity of the methods used and their sophistication. Selection criteria, comparability of incorporated studies, causality, and endogeneity are important aspects of study designs. In general, narrative reviews give a good historical overview of the existing literature, and theoretical papers point out interesting details of the CSP-CFP relationship. Vote-count studies use a simple and comprehensible but unreliable and potential misleading approach to test the linkage. Finally, meta-analyses are the state-of-the-art studies; they consider and shed light on all the qualitatively important aspects of an elaborate study. The findings of all reviewed meta-analyses show significant positive correlations between CSP and CFP and even better, a causal relationship. Although strongly debated for many decades, these findings, based on a sophisticated methodology probably demonstrate an existing positive causal effect of CSP on CFP.

The empirical stock market study directly addresses the research question, whether it is possible to achieve a superior return by pursuing a SRI strategy, and finds divergent results. The equal-weighted portfolio produced a significant risk-adjusted underperformance of -4.1% over a period of three years (April 2013-April 2016) compared to the S&P 500. In contrast, the value-weighted portfolio yielded an insignificant annualized alpha of 0.46%. The implications of these findings are limited due to a variety of the parameters chosen. A superior return cannot be concluded from the SRI strategy implemented here.

Table of Contents

List of Figures

List of Tables

List of Formulas

List of Abbreviations

APT	Arbitrage Pricing Theory
AuM	Assets under Management
CFP	Corporate Financial Performance
CSP	Corporate Social Performance
CSR	Corporate Social Responsibility
EPS	Earnings Per Share
ESG Factors	Environmental, Social, and Governance Factors
FNG	Forum Nachhaltige Geldanlage
GSIA	Global Sustainable Investment Alliance
HML Factor	High Minus Low Factor
ILA	Influential Literature Analysis
ILO	International Labor Organization
MPT	Modern Portfolio Theory
MRP	Market Risk Premium
OECD	Organization for Economic Co-Operation and Development
ROA	Return On Assets
ROE	Return On Equity
SMB Factor	Small Minus Big Factor
SR	Sharpe Ratio
SRI	Socially Responsible Investing
UNPRI	United Nations Principles of Responsible Investing
WBCSD	World Business Council for Sustainable Development

Chapter 1 Introduction

In recent years, socially responsible investing (SRI) has grown steadily among both institutional and private investors. Current figures from the Global Sustainable Investment Alliance (GSIA) show the development of the last few years distinctly: the total global assets under management[1] (AuM) of sustainable invested money increased between 2012 and 2014 from $ 13.3 trillion to $ 21.4 trillion, an increase of 61% in two years (GSIA, 2015). This development is visible at both an international level and a national one. To give an example, the SRI market in Austria, Germany, and Switzerland grew fifteen-fold between 2005 and 2014 (FNG, 2015). The importance of this specific investment approach can be seen from its share of total AuM in the US. In 2014, 18% of the total American AuM was managed with respect to sustainable considerations; this is not a negligible amount (GSIA, 2015).

1.1 Objective and Procedure

Given these facts, questions arise: What was or is the driver of this rapid development? Why do investors honor firms that pursue social goals besides the traditional objective of maximizing shareholder value? Is there a possibility to achieve higher returns on such socially responsible investments or do these investors simply have altruistic preferences? This study examines the first motivation (abnormal return) and seeks to determine whether this kind of motivation can be empirically justified, specifically if it is really possible to beat the market by implementing an SRI strategy. An implied precondition for such an over-return is a positive relationship between the sustainable conduct of a firm and its financial performance, referred to as the CSP-CFP link in the academic literature. For that reason, the study also examines this interesting relationship. This is done in two ways. A review of the past 50 years of the published literature focusing on the CSP-CFP link gives important insights into this interesting relationship. In the second part of the paper, an empirical study implementing an SRI investment strategy furnishes further evidence for a positive CSP-CFP relationship. Considering the vast volume of existing studies on this topic, one can ask why there is a need for another study. The answer is twofold. For one, precisely because of this gigantic number of studies, there is a need to summarize the status quo. Secondly, the data applied to the empirical study are quite new and could provide an information advantage.

---

[1] Assets under Management (AuM): total of financial assets which a third party (e.g. fund manager, private bank etc.) manages on behalf of the owner.

The structure of this paper is as follows. Chapter 1 presents an introductory overview of the topic of SRI with its relevant components. The chapter shows general definitions, developments over time, hypotheses concerning the direction of the CSP-CFP link, empirical problems, and CSP and CFP measures. The idea is to assist the reader in understanding the upcoming content and to explain why the subsequent literature review is necessary. Chapter 2 contains the literature review, which gives the insights mentioned above into the current stage of academic research. Important aspects of the existing literature are then used to construct the empirical study. Chapter 3 explains and discusses the empirical study with its design and results. To round things up, Chapter 4 comprises a brief conclusion.

1.2 Definition of SRI

A great variety of definitions of SRI can be found in both the academic literature and in practice. It should also be noted that that there are many similar investment approaches, such as *responsible investing*, *sustainable investing*, *impact investing*, and others, which in some cases are scarcely distinguishable from each other. Nevertheless, two definitions are frequently used. The first is the definition given by the Global Sustainable Investment Alliance[2] (GSIA). The GSIA denotes all such investment approaches with the expression *sustainable investing (SRI),* by which it means an investment approach that considers environmental, social, and governance (ESG) factors in portfolio selection and management. This approach encompasses a range of strategies: negative screening, positive screening/best-in-class screening, norms-based screening, integration of ESG factors, sustainability-themed investing, impact/community investing, and corporate engagement/shareholder action (GSIA, 2015). The individual strategies are explained below (see pages 14-15).

The second definition is well known as the Darmstadt Definition of Sustainable Investments. This definition of sustainable investment encompasses three aspects: economic, ecological, and social/cultural aspects. Each aspect must be satisfied by sustainable investments in stocks, bonds, funds, or direct investments. The *economic perspective* requires a mainly long-term orientation in generating profits in a legal way instead of short-term profit maximization by doubtful means. From the *ecological perspective*, sustainable investments should have a positive ecological impact (e.g. increase of resource productivity, recycling). Finally,

---

[2] Global Sustainable Investment Alliance (GSIA): association of regional sustainable investment forums, such as the European Sustainable Investment Forum (EURSIF) or the American Sustainable Investment Forum (US SIF). The regional SIFs are associations of national institutions, such as the Swiss Sustainable Finance (SSF). All these institutions have the mission to support sustainable finance (Eurosif, 2016).

sustainable investment profit making must be in line with the development of human capital (good employment conditions, education etc.), social capital (no discrimination, commitment towards responsible citizenship etc.), and cultural capital (respect and empowerment for cultural diversity etc.) to satisfy the requirements of the *social perspective*. Hence, Sustainable investing goes hand in hand with sustainable development (Hoffmann, Scherhorn, & Busch, 2004).

Another term often encountered in the context of sustainable investing and ESG factors are the United Nations Principles of Responsible Investing (PRI). These principles have been an important driver of ESG investing in recent years (Fulton, Kahn, & Sharples, 2012). Subscribers to these six principles commit themselves to abide by them, which both guarantees a sustainable investment process and behavior on the part of institutional investors and accelerates the SRI dispersal further. The six principles comprise the commitment to incorporate ESG issues in investment analysis and decision-making processes, to be active owners and incorporate ESG issues into ownership policies and -practices, to seek appropriate disclosure of ESG issues by entities that the investors invest in, to further promote the acceptance and implementation of these principles within the investment industry, to work together with other investors to enhance efficiency, and to report all activities and efforts towards the implementation of these principles (PRI, 2016).

In this paper, the first definition, from the GSIA, is used in the same broad sense and with its incorporation of several related investment approaches. However, two crucial distinctions are made here. Firstly, *impact investing*[3] and *philanthropic investing*[4] are not considered in the following parts of the paper. Second, it is assumed that socially responsible investors have a primary goal of maximizing their return. Another important point is that SRI is a rather new concept. The novelty of the SRI term is the reason why it cannot be found in older research papers. The implied coherence between CSP and CFP and its discussion has lasted for much longer. Bragdon Jr and Marlin (1972) reasoned about this topic over 40 years ago. Besides linguistic issues, there is indeed a difference between SRI in the narrow sense and the subject of most papers which have examined the CSP-CFP link. Although SRI implies a relationship between CSP and CFP, studies concerned with SRI (in general fund studies) only indirectly

---

[3] Impact investing: impact investing is a subset of SRI. It is an investment approach which not only considers ESG factors in the capital allocation, but explicitly wants to have a positive impact on the environment or community, too. Compared to philanthropy, the financial return builds the center of impact investing (Investopedia, 2016).

[4] Philanthropic investing: philanthropic investing is also a subset of SRI, but a rather extreme one, often referred to as donation. The main point of philanthropic investments is the social return (Investopedia, 2016).

address the CSP-CFP linkage when they compare the performance of SRI to that of conventional investments. In contrast, the majority of research papers address the CSP-CFP link directly. But at the end the question driving all studies is still whether sustainable firms or investments generate a higher financial performance. This is why this and the next chapter outline the current state of knowledge in this field broadly to reflect on the CSP-CFP relationship as a necessary condition for successful SRI. Hence, the study does not treat fund studies (SRI studies) differently or separately. Chapter 3 then focuses more specifically on SRI, when the returns of a SRI strategy are examined. But before that, another concept that often appears with and is related to SRI is explained and delimited in the next section.

1.3 Corporate Social Responsibility (CSR) and SRI

In academic literature, the two concepts CSR and SRI are often used in the same context and mixed with each other. And indeed, a clear segregation is hard to define. Nevertheless, a distinction is made here. The World Business Council for Sustainable development (WBCSD) defines CSR as follows: "Corporate social responsibility is the continuing commitment by business to behave ethically and contribute to economic development while improving the quality of life of the workforce and their families as well as of the local community and society at large." (World Business Council for Sustainable Development, 1999, p. 3). Corporate social responsibility (CSR) is a concept which originates from management theory and focuses on consequences of managerial decisions on environmental, social and governance matters (World Business Council for Sustainable Development, 1999). It can be seen as a concept which aims more at the interior of a firm, which suggests moral principles that managers should implement to satisfy stakeholders' demands, whereas SRI views the firm from the outside. SRI stems from the field of finance and, as mentioned in the preliminary passage, as an investment approach, it aims at investors rather than at firms and their managers. Although scholars often claim to examine the relationship between CSR and financial performance, they often treat CSR as a synonym for CSP and their procedure resembles that of proving a positive CSP-CFP linkage, or indirectly that of finding evidence for the advantage of socially responsible investing (Cochran & Wood, 1984). For that reason, research papers concerning CSR related to the CSP-CFP link are also considered in the following literature review.

1.4 Market for SRI and Development

As stated at the beginning of this chapter, the market for sustainable investing has increased significantly over the past few years. $ 21.4 trillion AuM account for 30% of the total AuM in the GSIA regions: USA, Canada, Europe, Japan, Asia, Australasia, and Africa. The US, Canada, and Europe grew most rapidly. The dominance of these three regions is also evident when looking at the total volume of sustainable investments: combined, they account for 99% of all sustainable investments (GSIA, 2015). Not only has the volume of sustainable investments increased; the segment has developed, visible in the structure of investors and in the investment strategies evolved and offered. GSIA reports that the global proportion of institutional and retail investors had shifted during the period of 2012-2014 in favor of the latter (2012: 89%/11% vs. 2014: 87%/13%) (GSIA, 2015). Greater transparency and new measurements will raise retail's share further. In Europe, the distribution of the different investment classes also changed. In 2011, stocks accounted for 33% of total allocated funds compared to 50% two years later, a spectacular raise. In the same period, bonds lost 11% and notes 6% (Eurosif, 2014). According to the most recent report of FNG[5], new investment strategies have been introduced in the field of sustainable investments in the last 10 years. Examples include *norms-based screening, impact investing, ESG integration* and *theme-based investing* (FNG, 2015). The next sections briefly describe today's most important strategies.

Negative Screening

Negative screening is the simplest and hence oldest sustainable investment strategy. In particular cases it is followed automatically by obeying international law, which for instance prohibits investments in anti-personnel landmines. More generally, this strategy excludes so-called *sin stocks* (tobacco industry, adult industry, gambling, arms, and others) from the investment universe. SRI funds often use a negative screening approach as a first step of their asset allocation (Eurosif, 2014).

Positive Screening

This strategy consciously chooses assets which perform very well regarding a specific sustainability measure (e.g. ESG factor) (Eurosif, 2014).

---

[5] FNG (Forum Nachhaltige Geldanlagen): professional association of SRI. Its members consist of banks, asset managers, insurance companies, NGO's and others (FNG, 2016).

Best-in-Class Screening

Best-in-class screening is a subcategory of positive screening. The investor picks the relative best company regarding a specific measure (e.g. ESG factor) from every sector. This strategy has the advantage that all sectors are involved in the process, and no sector is excluded because of its sector-specific characteristics (e.g. higher pollution in the oil/energy industry). This approach is the most commonly used, especially by mutual funds (Eurosif, 2014). After SRI funds have excluded unethical sectors by negative screening, they usually use the best-in-class method to optimize their portfolio

Norms-based Screening

This screening approach is based on norms and standards defined by international organizations such as the OECD, ILO, UN, and others. This can result in an exclusion of investments which are not in line with standards and norms (GSIA, 2015).

Theme-based Investing

This strategy focuses on specific sustainable topics such as food, renewable energy, climate change or agriculture (GSIA, 2015).

ESG Integration

This strategy explicitly considers environmental, social, and governance (ESG) factors in the capital allocation. This strategy can be combined with a best-in-class strategy (GSIA, 2015).

Impact Investing

Impact Investing is a subset of SRI. It is an investment approach which not only considers ESG factors in the capital allocation but also explicitly wants to have a positive impact on the environment or community. Compared to philanthropy, the financial return builds the center of impact investing (Investopedia, 2016).

Engagement and Shareholder Activism

This strategy involves investors' action to have an impact on companies' behavior. A precondition is a sufficient voting power (equity share) or organization of this power among private shareholders. Direct engagement (communication with the board/ management), filing of shareholder proposals and proxy voting are types of shareholder activism (GSIA, 2015).

1.5 Theoretical Foundations of the SRI Debate

As defined above, SRI is an investment style which incorporates ESG aspects in the investment process. The underlying assumption of this style is that *corporate social performance* (CSP) is positively correlated with *corporate financial performance* (CFP). Hence, if the CSP-CFP link is positive, SRI pays off. That is why this study examines this key relationship more closely in Chapter 2. Before the existing literature and their contributions are discussed, a short summary of the basic positions and their supporting theoretical arguments within the SRI debate are given here. Of course, this summary contains findings of the existing literature, but an overview is given here to help the reader to obtain a better understanding of the topic.

The substantial question whether CSP leads to better CFP, and hence whether it can be achieved a higher return by investing in such sustainable assets, has been debated for many decades. Basically, there are two traditional views on this matter: the shareholder view and the broader stakeholder view. In 1962, Milton Friedman published a fundamental *shareholder theory*, which has been quoted several times in various contexts since then. The essence of this theory is that the main function of a firm's management is to maximize the firm value on behalf of its shareholders. Any other usage of corporate assets (e.g. donations, environmental preservation and others) equals harm to the interests of the company's shareholders, since it decreases shareholders' profits (Friedman, 1962). Hence, a company should not pursue socially responsible behavior. The shareholders themselves are the individuals who can finally decide whether they want to be socially responsible or not (Friedman, 1970). In contrast to this, another influential theory was formulated by R. Edward Freeman in 1984, known as *stakeholder theory*. Freeman defined stakeholders as additional groups (suppliers, employees, customers, the state and others) with claims on the firm besides the shareholders, hitherto considered central. In this perspective, a company has a responsibility not only towards its shareholders but even more towards its environment and other stakeholders. Thus, it can make sense for the management and ultimately for the shareholders when the firm spends money on social and environmental activities if environmental or social costs can be avoided (maximizing firm value). In this view, the firm value of socially responsible firms will be higher than that of their counterparts.

These two positions are so important in supporting specific opinions on the value of corporate social behavior that they are mentioned in almost every paper in this field (Preston &

O'bannon, 1997; Ullmann, 1985; Waddock & Graves, 1997). Examining SRI investments can indicate theoretical reasons for and against SRI as well. On the one hand, supporters argue that there is a social risk premium for which the market has to compensate. The long period over which positive results have been found empirically can be used as evidence (Ambec & Lanoie, 2008; Clark, Feiner, & Viehs, 2015; Cochran & Wood, 1984). Others explain (risk-adjusted) superior returns with information inefficiency, as do Derwall, Koedijk and Ter Horst (2011) in their study. The three authors reason that a superior return results because investors underestimate the performance of sustainable companies compared to their counterparts, which results in lower stock prices. These companies then cause positive surprise by generating higher than expected returns (*error-in-expectation hypothesis*). On the other hand, opponents contend that the SRI selection process with its restrictions contradicts modern portfolio theory, which makes for lower diversification and accordingly lowers the efficient frontier, which finally causes the investor's return to suffer (Le Maux & Le Saout, 2004; Markowitz, 1952). Higher information and transaction costs compared to conventional investments are also indicated, especially if the performance of funds is compared (Barnett & Salomon, 2006; Derwall, Guenster, Bauer, & Koedijk, 2005).

Besides these theoretical perspectives, the linkage of CSP and CFP has been empirically examined. Preston and O'Bannon (1997) summarize six prevailing hypotheses of different causal CSP-CFP relationships. Additionally, McWilliams, Siegel and Wright (2006) assume a neutral link. The relationship shapes have been tested and proven numerous times by successive studies.

Social Impact Hypothesis (positive CSP-CFP link)
This hypothesis corresponds with Freeman's stakeholder theory. Expenditures for social and environmental efforts result in a higher financial performance. Finding evidence for the social impact hypothesis is the intention of most studies (Salzmann, Ionescu-Somers, & Steger, 2005). Examples for studies which back the social impact hypothesis include Pava and Krausz (1996) and Preston and O'Bannon (1997).

Trade-Off Hypothesis (negative CSP-CFP link)
The trade-off hypothesis is in line with Milton Friedman's propositions. Social activities are connected to additional costs, which cannot be compensated by additional profits. Hence,

corporate social behavior lowers firm value and harm its shareholders (Salzmann et al., 2005). Evidence has been found by Vance (1975).

Financial Slack/Available Funds Hypothesis (positive CFP-CSP link)

The financial slack hypothesis (also termed the available funds hypothesis) specifies a reverse linkage. A good financial situation leads to good corporate social behavior, founded in the financial capability to do so. McGuire, Sundgren and Schneeweis (1988) support this hypothesis. This hypothesis shows that it is not enough to prove a positive correlation[6], but it is necessary to test the direction (nature of causality) of the relationship as well (Preston & O'Bannon, 1997).

Managerial Opportunism Hypothesis (negative CFP-CSP link)

Better financial performance has a negative impact on social and environmental projects. The reason for this is a principal-agent relationship between shareholders and the management. Available funds seduce managers to discretionary conduct (Preston & O'Bannon, 1997). Evidence for the managerial opportunism hypothesis was found by Posner and Schmidt (1992).

Positive/Negative Synergies

In the positive case, social impact and available funds coexist, meaning that the exact directional characteristic cannot be stated. Good management achieves good financial performance and cares about stakeholders, resulting in a positive synergy. In the negative case, managerial opportunism and trade-off coexist, which results in a poor performance of both CSP and CFP (Salzmann et al., 2005). Orlitzky, Schmidt, and Rynes (2003) found evidence for such a positive virtuous cycle which was first suggested by Waddock and Graves (1997).

Supply and Demand Theory (neutral link)

A neutral outcome is also a feasible empirical result. Theoretical support is provided by the supply and demand theory, which suggests that a firm performs social activities on the specific scale which is demanded by the market and other stakeholders. This level of social activities maximizes the firm's profits (Salzmann et al., 2005). Aupperle, Carroll and Hatfield

---

[6] Correlation: the degree to which two or more attributes or measurements on the same group of elements show a tendency to vary together (Dictionary.com, 2016). The correlation does not imply any causal relationship.

(1985) did not find a clear relationship in their study, which provides an example of a neutral link.

Other Links

Apart from these dominant descriptions of the CSP-CFP relationship, additional shapes have been discovered. A U-shape was found by Bowman and Haire (1975), explained with costly social investments which have a positive financial impact if they are large enough. According to this finding, a manager should either care nothing about the firm's environment or invest at least a threshold level in social issues. In contrast, Lankoski (2000) found evidence in his study for an inverted U-shape link between CSP and CFP. This implies that there exists an optimal level of CSP expenditures which maximizes financial performance. Firms which invest less still have potential to benefit from positive marginal utility (they can improve financial performance). Since CSP is expensive, overinvestment results in decreasing profits. To illustrate this, one can imagine a company that commits itself to yearly donating a sum of money for humanitarian projects. This improves the firm's reputation, which boosts its revenue. After reaching a certain volume of charitable contribution, the revenue will increase less than the donations do, which finally results in waning profits.

Among such a broad variety of results, it seems possible to find support for every kind of relationship, which could make a final answer impossible. Other authors have raised some arguments which make this ambiguity plausible and suggest further research in this field under exactly defined conditions (e.g. research within a particular industry) (Ullmann, 1985).

1.6 Empirical Problems

A recurring element of past literature reviews has been the criticism of preceding studies regarding conflicting results and hence the lack of a conclusive result for the CSP-CFP link (e.g. Waddock and Graves (1997) found positive evidence, Arlow and Gannon (1982) negative results and Aupperle et al. (1985) found neutral evidence). Reasons for these contradicting results are various. Thirty years ago, Ullmann (1985) lamented a lack in theory, inappropriate definition of key terms, and deficiencies in the data base. In the following decades, succeeding studies have criticized the same points, indicating that the problems may be of a deeper nature or even that they are not completely resolvable (Griffin & Mahon, 1997; Wagner, 2001).

The next section presents a short overview of the different measures employed by authors for CSP and CFP. This can help to understand the persistent difficulty of drawing clear conclusions from the great variety of papers.

Measures for CSP

Corporate social performance is defined as a multidimensional construct which involves all efforts of the company to satisfy the economic, legal, ethical, and discretionary (e.g. philanthropic) demands of its stakeholders (Carroll, 1979). Finding appropriate measures for the social performance of a company has been very challenging and still is. Thus, the usage of certain measures is fiercely debated by scholars. There are even authors who cast doubts on the legitimacy of CSP as a measurement for corporate behavior at all. Rowley and Berman (2000) argued that CSP is neither an empirically nor a theoretically viable construct (aggregation of unrelated variables) and thus cannot be used to prove a general CSP-CFP link. The number of studies which have tried to explain this link shows that most scholars do not share this opinion. The leverage points to measure sustainability (or CSP) of a firm are as multifaceted as breadth by which the definition of sustainability can be interpreted. Some authors use one specific aspect of sustainability (e.g. pollution) (Shane & Spicer, 1983), whereas others use a multidimensional approach (e.g. pollution, remuneration policies, working conditions) to measure the performance (Lerner & Fryxell, 1988). Moreover, all these aspects can be gauged in many ways. For instance, the level of pollution can be gauged by CO_2 emissions, NO_x emissions, or others. In addition, Schreck (2008) distinguishes between three sources of information: attitudes/perceptions (e.g. interviews of experts), performance figures (e.g. pollution), and (corporate) disclosures. An overlapping source is provided by scores of third parties, who evaluate the firm's efforts in environmental, social, and governance dimensions. Older studies are often based on reputation indices (a popular one was generated by Milton Moskowitz), which rate firms according to their social performance/disclosures, or content analysis on the basis of firm disclosures (Bowman & Haire, 1975; Cochran & Wood, 1984; Moskowitz, 1972). New studies often use ESG scores provided by third parties (Auer & Schuhmacher, 2016). Examples of score providing institutions are MSCI, Sustainalytics, and Thomson Reuters (MSCI (2016), Sustainalytics (2016), Thomson Reuters (2016)).

Measures for CFP

Strategies for measuring the financial performance of firms are similarly dispersed. Here, three classes of measurements can be differentiated: accounting-based, market-based, and perceptual (survey) measures (Orlitzky et al., 2003). Accounting ratings are calculated based on figures in the balance sheet and indicate the internal efficiency of a firm. Examples include return on equity (ROE), return on assets (ROA), earnings per share (EPS), and Tobin's Q (Choi & Wang, 2009). Market-based measurements include the stock performance of a firm, price-earnings ratios and others, which can differ from figures obtained through an accounting view (Albertini, 2013). Other studies use financial risk as an indicator of financial performance, which has implications for capital costs, firm value, and finally investors' return (Orlitzky & Benjamin, 2001). Market-based measurements indicate shareholders' satisfaction with the company's activities. Perceptual measurements represent opinions and estimations of market observers and are rather subjective (Orlitzky et al., 2003). In contrast to CSP, the procedures to generate CFP measures (e.g. ROA) are theoretically well defined. But which of the various CFP measurements best represents the financial performance of a firm remains an object of discussion. Consequently, it can be assumed that the lack of standardized measurements of both CSP and CFP, will always lead to different results among studies. A solution to this problem are meta-analyses, which use statistical tools to synthesize the results of different methodologically distinct single studies (Kurtz, 2005).

Chapter 2 Literature Review

This chapter addresses the current status of literature concerning the CSP-CFP link. The first section explains the goal and justification for a further literature review in this field. The next sections present the search process, the applied literature selection and categorization, and the evaluation criteria applied in the review. Then the different literature categories are discussed. This is followed by a final discussion of the findings across all categories.

2.1 Objective and Justification

The objective is to give an overview of studies with a focus on qualitative aspects. The rationale behind this idea is, firstly, that no review has done this so far and, secondly, there is a correlation between the quality of a paper and its results. The review focuses on secondary literature, thus reviews and meta-analyses of primary studies, for two reasons. First of all, the volume of primary studies is almost impossible to cover by a single study, although there are studies which claim this (Friede, Busch, & Bassen, 2015). The second reason is that the divergent results of primary studies are less expressive than aggregated outcomes of secondary papers. Accordingly, this study tries to generate more meaningful results. Similar meta-studies have often failed to assess primary studies on a qualitative level. They focused rather on the final outcome (positive/negative CSP-CFP link) than on the methodology and the consequent weight/quality of the result. Moreover, the flood of past studies shows that it is difficult, perhaps even impossible, to find a conclusive result for the CSP-CFP link. Margolis et al. (2003) mentioned an ongoing debate with no result: "The steady flow of research studies reflects ongoing efforts both to resolve the tension between advocates and critics of corporate social performance and to shore up the methodological and theoretical weaknesses in past studies[…] The imperfect nature of these studies makes research on the link between CSP and CFP self-perpetuating: each successive study promises a definitive conclusion, while also revealing the inevitable inadequacies of empirically tackling the question. As the acceleration in the number of studies reveals, research that investigates the link between CSP and CFP shows no sign of abating." (Margolis & Walsh, 2003, p. 278). This study addresses this issue and examines the existing secondary literature based on its research approach. It does not seek a conclusive result, meaning a positive/negative CSP-CFP relationship, but a quality-controlled overview of the existing research, which simultaneously highlights important findings on primary or secondary research levels. This new research direction is one of the arguments which justify an additional review of this topic, but there are more. Endrikat,

Guenther, and Hoppe (2014) give another reason when they argue that the CSP-CFP link is time dependent, which motivates further primary research and has finally to be summarized in a review. In addition, no review has successfully incorporated the whole set of primary studies in this field so far (Endrikat et al., 2014). These three arguments explain further need for this qualitative review.

2.2 Literature Search

The literature search was made between the 1 April 2016 and the 30 April 2016. Existing literature was found through systematic search of well-established databases: *Google Scholar, Science Direct, Web of Science, SSRN, Ecobiz, Sage, NBER and EBSCO.* Applied keywords comprised *sustainable investing, socially (responsible) investing, SIR investing, ESG investing, CSP-CFP link, corporate social/financial performance,* and *CSR investing.* All these keywords were jointly utilized with the expressions *meta-analysis, meta-study,* and *(literature) review* to obtain especially the secondary papers in this field. Only freely available studies were considered, which naturally risks neglect of important findings, but it is presumed here that in the academic interest significant research findings should be publicly available without limitation. In a second phase, cross-citations of eventuated studies were also explored to improve search results. That procedure yielded a total of 73 secondary-level research studies.

2.3 Literature Selection

Once a whole set of potentially relevant research studies had been garnered, the contained papers were filtered further. The objective of this process was to focus firmly on the interesting topic of financial performance as a consequence of sustainable corporate behavior, so that sustainability has been covered as completely as possible. To achieve this goal, papers focusing only on one part of the ESG criteria were excluded (see Appendix 6.1 for details). This reduced the sample from 73 to 49 studies. In addition, studies whose main topic was not the CSP-CFP relationship or studies which focused too closely on the firm-internal view were also excluded (Aguinis & Glavas, 2012; Carroll & Shabana, 2010; McWilliams et al., 2006; Wood, 2010) Finally, 45 studies remained.

2.4 Literature Categorization

As explained in the introduction of this chapter, the main goal of this literature review is to give a quality insight into the existing literature on the CSP-CFP link and on socially responsible investing. Accordingly, one possible categorization of papers is applied which is considered reasonable here. *Figure 1* illustrates the literature selection procedure employed. The resulted papers are divided into four categories, which are explained in the following passage. Afterwards, the categories and the contained papers are discussed in detail.

Figure 1: Literature Selection and Categorization
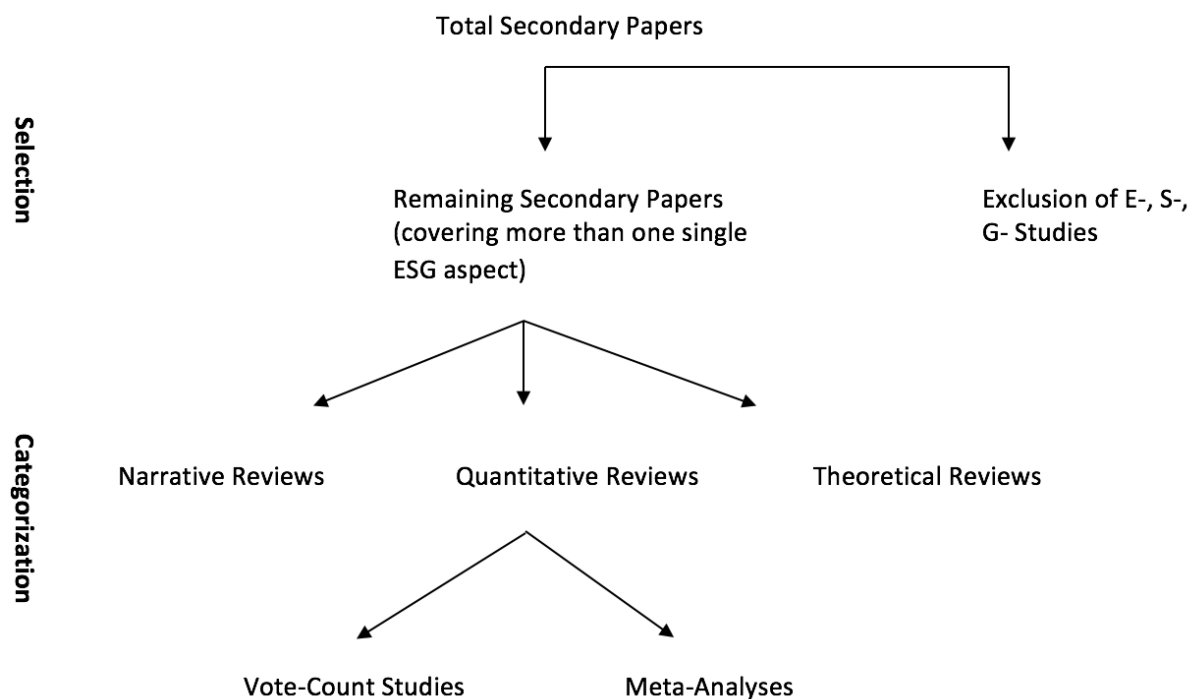
Source: Own Presentation

Endrikat et al. (2014) distinguish in their study between narrative approaches, vote-count studies, and meta-analytical approaches. A similar distinction is used here. Generally, *narrative reviews* and *quantitative reviews* are differentiated. Whereas narrative reviews summarize in a descriptive way, quantitative reviews do this by means of numbers. Quantitative reviews can be further differentiated into vote-count studies and meta-analyses. Vote-count studies put primary studies into positive, negative, and mixed/neutral result categories and assess the general CSP-CFP relationship based on the proportions of these three categories. This fairly simple approach has been object of dispute several times. Academics have criticized the fact that the vote-count technique could lead to biased results,

because it does not take into account differences in sample size, effect sizes or correlations. To illuminate the problem, one can imagine ten studies, four of them with a significant positive and six with non-significant results. Vote-counting would now suggest that there is no effect, since the significant positive results are in a minority. The problem is that such a conclusion can be fatally misleading, because there may be an effect, but in these other six studies the sample sizes were too small. To make matters worse, the power of this procedure decreases with increasing numbers of studies included (Hedges & Olkin, 1980). Even though these drawbacks have been known for decades, vote-count studies are still widely performed by researchers (Friede et al., 2015; Griffin & Mahon, 1997; Roman, Hayibor, & Agle, 1999). But there is a technique which can address this difficulty: meta-analyses. Various concepts are understood as meta-analyses. Beurden and Gössling (2008) define meta-analysis as an instrument to recapitalize prior results based on descriptive statistics. Another definition characterizes a meta-analysis as an instrument which synthesizes precedent findings and corrects for characteristics of the data sets included, such as sample sizes (Orlitzky et al., 2003). Meta-analyses in sense of the second definition are currently the most sophisticated instruments to synthesize multiple, inconclusive or conflicting results (Endrikat et al., 2014). The advantage of this technique is that it corrects for sampling errors and measurement errors of the integrated data sets of prior studies (Boaventura, Silva, & Bandeira-de-Mello, 2012). The capability of meta-analyses to synthesize vast studies which implement distinct methods and generate differing results illustrates this superiority well; T-test statistics, t, and effect sizes, d, can be transformed to correlations, r, which then can be aggregated to a conclusive result (Schmidt & Hunter, 2015). From a technical point of view, it seems convincing that meta-analyses are the most objective and advanced method. Besides these categories, an additional class of research papers can be defined as theoretical reviews. These qualitative studies do not primarily aim at generating a conclusive result for the CSP-CFP relationship, but have the goal of contributing insights for future research. Put another way, they summarize preliminary work in a narrative style and try to improve upcoming studies.

2.5 Evaluation Criteria

In general, studies are evaluated based on qualitative aspects. Points of interest of secondary papers are their *employed methodology, technique to make the methodological distinct primary studies comparable, comprehensibility of selection process, consideration of sample*

sizes, size effects, significance of results and endogeneity[7]*, distinction between correlation and causality* and *comprehensibility of conclusion.* It is not easy to compare the quality of these papers across categories in every single point due to their different approaches. For example, theoretical papers give important insights into the current state of knowledge from a theoretical standpoint, but they do not quantify the CSP-CFP relationship. Hence, these papers cannot be judged based on their statistical sophistication as vote-count studies can be. This point justifies the categorization and the comparison of studies within these categories. Nevertheless, there are aspects which can be compared across categories, and these are the subject of the overall discussion at the end of the chapter.

Before the categories are discussed, it has to be mentioned that this categorization is not immune to subjectivity. In particular, the classification between narrative and vote-count studies can be debated since the borders are fluid and most narrative reviews draw their deduction of the CSP-CFP relationship by implementing a vote-count technique. Simultaneously, most vote-count studies summarize their findings in a narrative way. Nonetheless, it is considered reasonable here to make a distinction between these two kinds of studies and to classify the papers according to their main goal: Narrative reviews give a (historical) overview starting with the first published study and finishing at the end of the study's observation period; vote-count studies summarize all applicable studies and sum the results to obtain the proportions on which their conclusions are based.

2.6 Literature Discussion

This sub-chapter's objective is to discuss the different literature categories and to discuss the most important findings across all studies. Details of single studies are available in Appendix 6.2.

2.6.1 Narrative Reviews

Fifteen studies are classified as narrative reviews by virtue of their method and main purpose. The studies were investigated and compared based on their explanation of the literature selection process and methodology, their considerations of causality and endogeneity, and their conclusions. Authors often use narrative reviews as a first step to their own empirical

---

[7] Endogeneity: in the subsequent chapter endogeneity indicates an omitted variable bias, meaning that the ignorance of a certain factor can have distorting effects on the interesting relationship (omitted variable has causal effects). It is clear that endogeneity is related to causality but in this study causality defines the direction of relationship (CSP->CFP, CFP->CSP or concurrent) as implemented by prior studies (Orlitzky, Schmidt, & Rynes, 2003). Hence, the two subjects will be treated separately.

study (Cochran & Wood, 1984; Derwall et al., 2011; Huppé, 2011). Narrative reviews summarize existing studies and their outcomes, but they pay little attention to the final effect sizes. Since the methodological approach does not really need explanation, all studies only mention their purpose in the abstract/introduction. A negative point is that only a minor number of studies show their literature selection criteria (UNEP Finance Initiative, 2006, UNEP Finance Initiative, 2007). The absence of a clear selection procedure makes the whole study harder to interpret. Perhaps a particular result arises only due to a lack of particular data. This can only be recognized if the data are transparent. In this regard, an extraordinary selection method which is unique among all reviewed studies should be highlighted here. Hoepner and McMillan (2009) select primary studies with a method called influential literature analysis (ILA), which is based on citations, a common measure for the impact of a study in science. The method is thoroughly documented by the authors and hence comprehensible for the reader.

Usually, narrative reviews give a historical overview of the existing findings when they put various distinct studies together. This is not a problem as long as the study does not conclude a final CSP-CFP result. Scholars have shown that it is not possible to compare all results with each other due to moderating effects (Grewatsch & Kleindienst, 2015). Nevertheless, there are studies which try to do this and obtain inconclusive results (Aupperle et al., 1985). More though-out studies consider endogeneity in their study and categorize the studies they include according to the CSP/CFP measure employed, and even according to regions and industries (Ahmed & Uchida, 2009; Clark & Viehs, 2014; Huppé, 2011). Cochran and Wood (1984) were concerned about endogeneity in the CSP-CFP relationship already many years ago. That is why they incorporated in the second part of their work the variables asset age and asset-turnover to explain the link; they obtained significant results. What is more, the CSP-CFP relation remained significantly positive. In general, it can be assumed that studies that are aware of moderating/mediating effects generate more interesting results, because they examine the CSP-CFP relationship in more detail and try to find the plain relation between CSP and CFP. In the review, only eight out of fifteen narrative studies take moderating effects into account; three of them check for it. A very important differentiation has to be made between correlation and causality, which was already suggested by Aldag and Bartol (1978). But only one third of the authors of narrative reviews comment on that distinction in their work. Some of them mention that a distinction should be made, but it is not really clear in how far this insight has affected their study (Ahmed & Uchida, 2009; Schröder, 2014). Clark

and Viehs (2014) claim that causality was not verifiable based on their employed data. In contrast, other authors devote themselves to this matter and summarize causality findings explicitly or even conduct their own empirical studies (Huppé, 2011; Renneboog, Ter Horst, & Zhang, 2008). Not all conclusions of narrative papers are entirely reasonable. A summary of the most important findings and suggestions for further research would be a sound way to end a narrative overview. Raar (2001) shows how hard it is to deduce a final answer to the CSP-CFP relationship based on a narrative overview: *"There are no precise outcomes to this paper."* (Raar, 2001, p. 228). But in contrast, most studies try to give a final answer to the shape of the CSP-CFP relationship. Most of these conclusions are naïve, since they arise from a blind aggregation of different methodologically distinct studies, often without consideration of sample sizes or significance. It is also inadvisable to derive general results from a bundle of case studies, since their outcomes do not necessarily have implications for other regions, industries, or firms. Although the ILA literature selection process mentioned above is unique and probably superior, the conclusion derived from the related study sample is still debatable. Such a conclusion is deduced from a few impactful but still single studies. It is unclear which part overweighs: the reliability of impactful results or the fact that single results are hard to generalize.

All in all, the narrative studies reviewed here give a good overview of the existing literature, which is basically their main purpose. Many of them are aware of the endogeneity and causality aspects of the CSP-CFP link. But the conclusions have to be treated with caution.

2.6.2 Theoretical Reviews

The seven theoretical reviews were all published in the new century except one, which was produced in 1999. This is hardly a coincidence. Having in mind that Ullmann (1985) mentioned the need for a general theory and many successive authors who had found inconclusive results called for the same, the advent of a new research direction is no surprise. These studies share the aim of understanding why preceding results have been inconclusive, what has gone wrong in past studies, and how future research in this area could be improved. All theoretical studies explain their method in a fairly comprehensible manner, and naturally they consider endogeneity and causality in their reflections, since their main goal is to examine the CSP-CFP relationship on a theoretical level. Harrison and Freeman (1999) is the only study which does not consider causality or endogeneity at all, and its conclusion is very focused on three specific topics. Rowley and Berman (2000), who criticize the CSP construct

in general, distinguish between the different research designs of the studies they analyze. Three studies raise interesting explanations of moderating and mediating effects between CSP and CFP which can be used to explain causality and simultaneously check for endogeneity (Grewatsch & Kleindienst, 2015; Peloza, 2009; Perrini, Russo, Tencati, & Vurro, 2012). Grewatsch and Kleindienst (2015) explicitly analyze every single primary study for moderators, mediators, and control variables to conclude a causal chain from CSP to CFP. Similarly, Peloza (2009) classifies studies according to CSP and CFP on different levels, which is a new approach. He differentiates between *mediating metrics, intermediate outcome metrics,* and *end outcome metrics*. Based on the findings of the studies he differentiates, Peloza generates a causal framework. Unfortunately, it is not clear to what extent Peloza uses reliable results or whether he considers sample sizes or other descriptive statistics of the implemented literature. Another interesting approach to shed light on the causes of prior mixed results of fund studies is used by Rathner (2012). His empirical regression is understandable, and the study checks for endogeneity and sample sizes of primary studies, which makes the results more reliable. Finally, Salzmann et al. (2005) gives a broad overview of the current stage of literature, theoretical frameworks, and tools. This study classifies literature differently to all other studies reviewed here, namely according to its empirical or theoretical nature. All in all, the conclusions and results of these theoretical reviews are very interesting, because they give new insights and are deduced from elaborate methodologies.

These theoretical reviews give no conclusive evidence for a positive CSP-CFP link, but they can help to understand why and when such a linkage can exist and why prior empirical studies have failed to find conclusive evidence for that linkage.

2.6.3 Vote-Count Reviews

It has already been observed that the vote-count method is problematic when trying to draw conclusions from it. Nevertheless, this category contains fourteen papers, which comprises almost one third of papers reviewed in this study. Surprisingly, the majority of these papers were published in the last 15 years, even though the methodological drawbacks have been known for over 20 years. The reasons for this are probably the simple procedure and the easy comprehensibility for the readers, what unfortunately does not make conclusions more reliable. Nine papers out of fourteen disclose how they had searched for studies and how they had selected them. As mentioned, a transparent selection process helps greatly in interpreting the final conclusion. Most authors explain in their work how they categorized studies and

what they checked for, but hardly any author explicitly explains that he applied a vote-count approach to generate final results and why he did so. Obviously, vote-counting has been the standard in this research field for so long that its application no longer has to be justified. Exceptions are Revelli and Viviani (2013), who explain their methodological decision in a reasonable way. To make the variety of primary studies comparable, the majority of studies categorize primary studies according to the CSP or CFP measure implemented. A few of them also check for endogeneity, or are at least aware of it (Beurden & Gössling, 2008; Boaventura et al., 2012; Fulton et al., 2012; Revelli & Viviani, 2013; Wallis & Klein, 2014). It has to be underlined that only four out of fourteen studies show awareness of the distinction between correlation and causality (Boaventura et al., 2012; Dam, 2008; Fulton et al., 2012; Margolis & Walsh, 2003; Ullmann, 1985). Not one of all them explicitly examines the causal relationship between CSP and CFP. From a methodological point of view, these vote-count studies all do the same thing. They select primary studies and calculate the proportions of positive, neutral/insignificant, and negative results. One shows weakness in the classification of positive/negative results (Griffin & Mahon, 1997). The authors define negative CSP combined with negative CFP as a negative correlation, though it should be the other way around. This mistake has been corrected by a later study (Roman et al., 1999). It is conspicuous that even though a lot of studies differentiate their studies on characteristics such as the CFP measure used, almost all draw the final conclusion from all studies. Combined with the fact that hardly any study considers causality, it seems that the majority of researchers blindly sum correlations and deduce final conclusions from these. The fact that results from all kind of studies, using various different setups, not controlled for sample sizes etc., find access into the final assessment of the CSP-CFP linkage lowers the reliability of these conclusions significantly. A result of that is that conclusions often sound too short-sighted (Sjöström, 2011). But there are also some exceptions, which critically reflect their results or put them in relation to their study's limits (Clark et al., 2015; Dam, 2008; Fulton et al., 2012; Raza, Ilyas, Rauf, & Qamar, 2012). Fulton et al. (2012) for example discriminate underlying studies based on their character (CSR, SRI, ESG studies) and on single E, S, and G dimensions. Finally, they deduce results for every category, which makes more sense, since they compare similar studies. But even then, the results have to be considered with some caution, because the number of papers per category is very small, which makes results prone to biases.

Even though some of the reviewed vote-count studies incorporate reasonable considerations addressing and quantifying the CSP-CFP relationship, they fail to compare homogeneous results and subsequently to deduce reliable conclusions.

2.6.4 Meta-Analyses

The literature search process found nine meta-analysis studies. They were all edited in the last 15 years. One of them probably brought up the most impactful results since Ullmann in this field and so won the Moskowitz price for socially responsible investing (Orlitzky et al., 2003; Ullmann, 1985). But it is not only this study which generated meaningful results. It was followed by a number of very sophisticated and thorough papers which examined the CSP-CFP relationship in groundbreaking and detailed ways. Besides the superior research technique (adjustments for sample sizes etc.), the studies investigate endogeneity, prove causality, and look for differences across regions and time periods. Furthermore, they quantify the size effect of CSP on CFP and thus indicate the strength of this linkage. Despite all comments about inconclusive results in this field, meta-analysis results seem to agree on a positive CSP-CFP link. All nine meta-analysis suggest positive correlations, and with regard to their statistical sophistication, this could indicate a true positive but weak (average correlation is around 0.1) relationship between a firm's social performance and its financial performance.

In the quality evaluation of these papers, it emerges that all studies explain their literature selection criteria and method in detail. Except two, all studies discriminate the underlying studies based on CSP and CFP measures. The studies then aggregate the primary studies' size effects to a summarized size effect. Some authors calculate this size effect of all possible combinations of CSP and CFP, which gives interesting insights into the different strength of certain relationships (Margolis, Elfenbein, & Walsh, 2007, 2009; Orlitzky et al., 2003; Wang, Dou, & Jia, 2015). To shed light on the same issue, Allouche and Laroche (2005) are the first authors in this field to use a meta-regression approach. Aside from that, seven out of nine papers investigate causality in detail. It deserves mentioning here that Margolis et al. (2007) deal explicitly with primary studies which comprise all three possible directional relations (from CSP to CFP, from CFP to CSP and concurrent) and find evidence for the virtuous cycle suggested by Waddock and Graves (1997). Contrasting this outcome, a very recent state-of-the-art study only finds evidence for a one-directional relationship (from CSP to CFP) (Wang et al., 2015). But the important point here is that these studies not only look for correlations

22

but also want to know what causes what. Authors of these meta-studies are aware of endogeneity. Some of them use statistically advanced methods to check for moderating effects of certain variables. For example, Allouche and Laroche (2005) use a multivariate-regression method to identify moderators. An interesting result in this context is the study by Margolis et al. (2009), which finds no effect of control variables. This would imply that CSP had a plain effect on CFP not affected by any other variable.

The conclusions of the papers are comprehensible, since they are drawn from reasonable and sophisticated research designs. What makes the results even more reliable is the variety of statistical checks implemented. All authors use file drawer analysis[8] or meta-significance testing to avoid publication biases[9], an aspect of meta-analyses that is often criticized.

Friede et al. (2015) has to be mentioned as a special case in this literature review. This study, which was published very recently, is a review of secondary papers. The data, all available data from underlying primary studies paired with a meta-analytical approach, makes it possible to summarize a size effect with a groundbreaking expressiveness. Moreover, this study is the first that generates size effects based on vote-count results. Surprisingly, the outcome is very similar to the correlation figure of the meta-analysis. The authors also examine differences between regions, E, S, and G dimensions, and portfolio- and non-portfolio studies. As an overall summary of the CSP-CFP relationship, this study yields probably the most interesting result so far.

To summarize, meta-analyses are the most advanced methods with which to evaluate the CSP-CFP relationship. The differences between the studies are small (e.g. number of implemented studies), since they all have very sophisticated research designs which consider endogeneity, causality, and possible biases.

2.6.5 Discussion

This passage discusses important findings of the literature review across all categories.

It is conspicuous that only half of the studies reviewed disclose their literature selection process, and even then they mention search criteria but seldom exclusion criteria. Exceptions

---

[8] file drawer analysis: procedure which computes number of unpublished studies needed to affect the stated effect size (Rosenthal, 1991).

[9] publication bias: assumption that published papers propose higher effect sizes than unpublished papers, which can lead to biased results (Rosenthal, 1991).

are authors such as Wang et al. (2015), who explicitly mention their exclusion of redundant data sets in the meta-analysis or Hoepner and McMillan (2009), who explain the selection process based on citations in every detail. Narrative reviews seldom disclose their method and their literature selection process explicitly, and neither do vote-count studies, even though this would be especially important for vote-count studies. This is because the studies incorporated, distinct in methodology, sample size, time period, and region, have a significant impact on the final result. But again, there are some positive deviations from the generality that explain what they do (Fulton et al., 2012; Revelli & Viviani, 2013; Roman et al., 1999). Meta-analyses are state-of-the-art in this regard. They disclose exactly what kind of studies they incorporate and which they expel. This does not surprise, since the quality of the data is key for such quantitative methods. Few theoretical studies disclose their selection procedure and methodology.

An important factor is how secondary studies try to make the diverse primary studies comparable. Most authors discriminate primary studies based on their employed CSP and/or CFP measures (44% of total reviewed studies), which is a reasonable approach. Unfortunately, many vote-count studies calculate results from the total of examined papers instead of differentiated results per category. In this regard, meta-analyses can demonstrate once more their competitive advantage, since they can synthesize results grounded on various approaches and still yield a reliable result. All the same, even meta-analyses distinguish between different CSP and CFP measures to compare the different correlation coefficients. These differing coefficients give evidence that specific CSP/CFP measures are in fact important moderators of the CSP-CFP relationship, which was also the object of theoretical research in recent years. In this regard, it is also debatable whether it is advisable to conclude results from a study sample covering 30 years. In fact, most studies consider the whole literature since the first publication in 1972. Only a few address this issue and examine studies of a more recent period to explore whether the CSP-CFP relationship has changed in the last decade (Boaventura et al., 2012; Wang et al., 2015). A special variant which deserves mentioning is that used by Margolis et al. (2009), which tests the long-term effect of CSP and, in a separate experiment, the effect based on studies between 1998-2007. According to that study, the correlation decreased compared to the correlation calculated for the longer period. This could be interpreted as a learning effect among investors. Moreover, hardly any study makes regional distinctions. Again, meta-analyses are the exceptions in this regard and

conduct experiments with studies covering different regions (Friede et al., 2015; Revelli & Viviani, 2015).

Half of the studies reviewed here show awareness of the difference between correlation and causality. This distinction is important if scholars want to explain the existing relationship between CSP-CFP and the final implications for investors. Meta-analyses show the best approaches to find a causal relationship.

The majority of all papers (60%) show awareness of endogeneity problems. Most papers search the underlying primary papers for moderators and mediators. Many vote-count studies show awareness in their report, but this has no effect on their final calculation, which is made by means of papers, which may incorporate moderators or not. Narrative reviews mention the possibility of moderating effects of size and other variables in their overview, too. Most meta-analyses analyze moderating effects systematically and produce powerful insights (Allouche & Laroche, 2005; Margolis et al., 2009). Beside, theoretical reviews present reasonable frameworks of moderating and mediating effects which often also explain causal directions (Grewatsch & Kleindienst, 2015; Peloza, 2009; Perrini et al., 2012).

The conclusions of the papers are not all very comprehensible. In particular, conclusions of vote-count studies are often unreliable, since they propose a positive relationship only because the majority of the underlying primary studies show a positive correlation. In particular, the results of a handful of studies appear suspect, since they are positive (sometimes 88% or even 100%) and the contexts of the studies seem commercial-driven (Clark & Viehs, 2014; Fulton et al., 2012; Mercer, Carpenter, & Wyman, 2009; UNEP Finance Initiative, 2006, UNEP Finance Initiative, 2007). With regard to research quality, meta-analyses and theoretical reviews generate the most reasonable and reliable conclusions. It has to be highlighted again that meta-analyses draw their conclusions not only from significant but also from insignificant study results by means of correcting for sample and measurement errors. This allows them to produce a result from a much larger result sample than vote-count studies, which cannot correct insignificant results.

In summary, narrative reviews cannot evaluate the existing CSP-CFP link, but they can give a good overview of the historical research with its highlights and problems in this field. Vote-count studies are clearly inferior in their technique compared to meta-analyses, and their

outcome has to be interpreted very critically. Theoretical reviews open new research directions to address certain aspects of the CSP-CFP relationship.

Chapter 3 Empirical Study

After an insight into the currently available literature has been given, further evidence for the relationship between CSP and CFP is generated by a new empirical study. Part One explains the research goal. Part Two gives reasons for an additional empirical experiment. The next part defines parameters and presents the data used. Part Four explains the experimental design, and the last part discusses the results, the limitations and makes suggestions for future research.

3.1 Research Question and Objective

The goal of this research study is to find evidence for a risk-adjusted superior return (positive alpha) of a particular socially responsible investment strategy in the US equity market. The evidence is produced under the market conditions which prevailed between 2013 and 2016. Such a superior return also indirectly but simultaneously supports a currently existing positive CSP-CFP link in the US financial market, since such a positive correlation is a precondition for the pursuit of superior return.

3.2 Research Justification

In the flood of already existing studies, the execution of this additional one can be justified by several arguments. First of all, the CSP-CFP nexus is historically dependent, meaning that the relationship changes over time and has to be re-examined from time to time. A second argument is that this study uses a fairly new data set provided by Sustainalytics, which has been fully available through Bloomberg only since 2014. It can be maintained that due to this data novelty the information has not been completely processed in the stock prices. This reasoning holds of course only under the assumption of weak market efficiency[10] (Malkiel & Fama, 1970). In the following, the experimental design and the data set employed are explained. After that, the results are presented and discussed.

---

[10] Weak market efficiency: the efficient market hypothesis (EFH) is a theoretical construct established by Eugene Fama in 1970. The EFH differentiates three levels of market efficiency: weak, medium and strong efficiency. In an increasing manner it is not possible to achieve a superior return by encompassing past, public available or even private information (Malkiel & Fama, 1970).

3.3 Parameters and Data

3.3.1 Parameters

CSP

CSP is here defined as the measured performance of the firm investigated in environmental, social, and governance aspects, known as ESG dimensions. This performance is estimated through the ESG overall score assigned by Sustainalytics, an independent financial research company and available on Bloomberg.

CFP

This study understands CFP in a market-based sense as the stock return of stocks.

3.3.2 Data

Market Data

The study uses stock data from the S&P 500 (Standard & Poor's 500), which is a broad market index covering the 500 largest companies within the American economy. The S&P 500 is a market-valued index and originally a price index. More recently, a performance variant has been established, called *the S&P 500 total return,* which takes account of dividends, but which will not be of further interest in this study since the study abstracts from dividends. There are two reasons for the choice of the S&P 500. First and foremost, the S&P 500 is a very famous index, which brings some advantages with it. Many market participants, such as investors, analysts, and banks, consider the development of this index very carefully. Accordingly, the data availability is very good. Besides this, the sizes of the enterprises included guarantee a reliable market liquidity. The second reason is that the necessary ESG scores are available for all S&P 500 companies. Additional to the S&P 500 data, another data set of market values are necessary for the risk-adjustment process explained later. This set contains the Small-Minus-Big- and High-Minus-Low premiums (SMB-/HML premiums), which are calculated weekly and kindly provided by Kenneth R. French (2016). The third type of market data used here are risk-free rates. The financial crisis of 2007/2008 damaged the widespread belief in completely risk-free rates. Nevertheless, the tough political engagement of the US government within the markets and the advanced recovery of the US economy give reason to consider US governmental bonds and notes as almost risk-free. Due to historically low interest rates in the last few years and a long-term perspective, this study uses weekly 10-year US-governmental bond rates as proxy for risk-free rates.

ESG Data

The study uses overall scores from Sustainalytics, which are available on Bloomberg. Sustainalytics rates firms in all three categories (environmental, social, governance) with respect to various single criteria. Examples of environmental criteria include CO_2 emissions, NO_x emissions, and water consumption. Social criteria are employee policies, employee satisfaction, donations, and others. And finally to illustrate governance matters, management compensation policies, and board election procedures can be mentioned here. All these single criteria are synthesized in an overall score at the end. This procedure has been a subject of debate among scholars. Rowley and Berman (2000) argue that overall scores yield misleading information because they aggregate single effects with weak or no correlation, which could produce a final effect of zero. Hence, it is advisable to focus on a single sustainability factor. In contrast to that, this study follows Waddock and Graves (1997) and others, who found a positive CSP-CFP relationship based on overall assessments. Besides that, this study seeks to incorporate sustainability as a preferably complete concept; hence, all three ESG dimensions are considered. It has to be added that, unfortunately, the exact coverage and calculation of these scores is not disclosed by Bloomberg/Sustainalytics for reasons of commercial secrecy, which makes the scores lacking in transparency. A notable detail is that Sustainalytics rates single firms against its peer group, which is why the overall scores are published as percentiles. Thus, scores can only be used to construct a rating, since they do not incorporate any absolute meaning. This work uses the latest data, as of April 2016. The data set of Sustainalytics is chosen because it is fairly newly available on Bloomberg, as mentioned above. Nonetheless, this data source bears one substantial disadvantage with it. The historical data availability is very weak and only goes back to 2014. But the aggregation criticism mentioned above turns here into an advantage: Since overall scores insist of various single criteria, they are robust against minor changes in firms' environment or firms' internal activities and policies. To obtain a longer observation period, it is assumed here that the ESG scores remain stable over three years, such that a three-year period (29.3.2013 until 25.3.2016) can be examined.

3.4 Experimental Design

The empirical design partially follows Derwall et al. (2005), who found evidence for excess returns of sustainable stocks. Moreover, this research employs a best-in-class investment strategy as well. Best-in-class investment styles are often claimed to contradict the idea of sustainability, since they allow investment in sectors which naturally perform poorly in

sustainability (e.g. oil & gas sector). This fact would argue for a negative screening approach, which generally bars certain sectors from the investment universe and is very common among fund managers and other institutional investors (Eurosif, 2014). But, as mentioned in Chapter 1.2, this paper does not suggest any altruistic or other social preferences of investors. Thus, it is assumed that investors want to maximize their profits according to the frequently used economic model of a *homo economicus*[11]. According to these assumptions, a best-in-class approach is most appropriate, since it does not preclude any sector or stock, such that the performance could suffer from. Subsequently, the research steps are presented from the construction of a best-in-class portfolio until the results.

Step 1: Construction of an Equity Portfolio

First, an equity portfolio employing a long-short strategy is constructed implementing ESG scores. This means that the S&P 500 is examined sector-wise. The best performing five and the worst five according to the ESG scores are selected from every sector. Afterwards the strategy goes long (buy) on the best five and goes short (short-sell) on the worst five assets in each sector. Firms with no score are excluded. In sectors that consist of less than ten firms, the sample is divided by two so that the better half contributes positively and the worse half negatively to the portfolio's return. Two portfolios are designed by implementation of the investment style: one equal-weighted and the other value-weighted, to check for sensitivity to portfolio-weighting effects. The weighted portfolio is rebalanced at the end of each investment year, in total twice.

Step 2: Calculation of Performance and Risk-Adjustment

Next, the weekly rate of return of these portfolios is calculated as a weighted average of return for three years, 29.3.2013 to 25.3.2016, or 156 weeks. This procedure resembles that used by Derwall et al. (2005) and contrasts with methods that consider dividends as part of a shareholder's return, which would be theoretically exact and was incorporated by other authors (Abbott & Monsen, 1979). This is due both to complexity and to the fact that it is questionable whether the inclusion of dividends would change the result. Modern Portfolio Theory (MPT), established by Harry Markowitz, further developed by William Sharpe and later by Eugene Fama and Kenneth French, require further risk considerations besides the return evaluation (Fama & French, 1993; Markowitz, 1952; Sharpe, 1964). According to MPT, high returns are generated on a cost, which is denoted by high risk. Thus, generated

---

[11] Homo economicus: economic actor, whose preferences are maximizing his own utility.

returns have to be risk-adjusted before they are expressive, meaning before they can be compared with other investments. This was also considered by other studies in this field (Alexander & Buchholz, 1978). There are several ways to risk-adjust returns. This study implements Fama and French's approach. The main points of the underlying idea are these: The Arbitrage Pricing Theory (APT) suggests that a return generated on an asset is determined through the risk-free rate and the compensation for diverse with the firm's business correlated risks (Bodie, Kane, & Marcus, 2014). The compensation of these risks is represented by premiums, which are calculated from representing portfolios. Sensitivity measures, called betas, determine how much a stock's price changes as a result of a risk factor change. By taking defined risks into account, a risk-adjusted return results. A risk-adjusted portfolio return can now be explained with this base equation:

$$r_P = E(r_P) + \beta_{1,P} * F_1 + \cdots + \beta_{n,P} * F_n + e_P$$

Formula 1: APT Portfolio-Return Equation based on Bodie et al. (2014)

r_p = ex-post observable, actual generated return

$E(r_p)$ = ex-ante required return

β_n = portfolio's sensitivity towards premium change

F_n = risk factor

e_P = portfolio's idiosyncratic return

The Fama-French approach used in this study, known as the *Fama-French three-factor model*, is an application of APT. Fama and French claim three relevant risk sources. These are the *market risk*, represented by the market premium/CAPM premium (market return - risk-free rate), the *size risk* (small firms are riskier than big firms, and this has to be compensated with a higher return), represented by the SMB premium, and finally the *value risk* (firms with a higher book-to-market value (value stocks) have a higher risk and this has to be compensated), represented by a HML premium (Fama & French, 1993). There have been numerous of models which are all based on the APT, use similar equations but add other risk factors as it did Carhart (1997) when he added a momentum factor. In this work it is assumed that the CAPM premium together with the Fama-French factors cover all relevant risk factors.

To find evidence for a risk-adjusted superior return (called alpha) arising from the SRI investment strategy, the study uses a (multi)-linear regression (OLS) method (Universität

Zürich, 2016). First, weekly portfolio excess returns are regressed on the market risk premium, which yields *Jensen's Alpha* (Jensen, 1968) . Then, the weekly portfolio excess returns are regressed on all three risk factors of the Fama-French approach, which yields *Fama's Alpha*. Finally, a one-sample t-test[12] shows whether the betas are statistically significant or not.

$$r_{p,t} - r_{f,t} = \alpha_{p,t} + \beta_{p,MRP} * MRP_t \left(+\beta_{p,SMB} * SMB_t + \beta_{p,HML} * HML_t\right) + e_{p,t}$$

Formula 2: CAPM and Fama-French Formula based on Fama and French (1993)

$r_{p,t}$	= generated return of portfolio in period t
$r_{f,t}$	= risk-free rate in period t
$\alpha_{p,t}$	= superior return in period t
$\beta_{p,n}$	= portfolio's sensitivity towards premium change
MRP_t	= CAPM premium ($r_{m,t}$-$r_{f,t}$)
SMB_t	= excess return of small and mid-caps against large-caps
HML_t	= excess return of value stocks (high book-to-market ratio) against growth stocks
$e_{p,t}$	= portfolio's idiosyncratic return in period t

If alpha is positive, the investment generates a risk-adjusted superior return, which means that the investment generates a higher return than the linked risk would justify (Jensen, 1968). In perfect markets, such market opportunities vanish immediately because of arbitrage activities (Bodie et al., 2014). However, positive alphas are exactly what thousands of fund managers and investors look for– and this study of course, too. It should be noted that searching for positive alphas only makes sense in imperfect markets, as it is assumed here, and that only a t-test can tell whether the positive result is really statistically significant and therefore expressive.

3.5 Results, Discussion and Limitations

The Resulted Portfolio:

The procedure explained above yields a portfolio containing 94 stocks arising from ten sectors. The list of sectors and stocks is available in Appendix 6.3.1. For practical reasons, a

---

[12] T-test: a one sample t-test is a statistical method to check the reliability (significance) of a statistical generated result (Statistics Solutions, 2016).

few adjustments had to be made. Three firms have not existed for three years, so they have been exchanged for firms from the same sector with the next worse/better ESG score. Expedia, Cablevision and Torchmark replaced News Corporation "A", News Corporation "B", and Synchrony Financial. Market capitalization data were not available for Coca-Cola Enterprises. Instead, Mondelez was included in the portfolio.

The Resulted Returns

This section presents the results. Table 1 compares descriptive statistics of the portfolios and the S&P 500.

Table 1: Descriptive Statistics of Portfolios

	Equal-Weighted Pf	Value-Weighted Pf	S&P 500
Number of stocks	94	94	500
Mean return p.a.	-2.30%	3.74%	12.37%
Geo-mean return p.a	-2.33%	3.49%	10.20%
Volatility p.a.	2.68%	6.96%	19.46%
Beta (CAPM)	-0.05	0.21	1
Sharpe Ratio	-1.72	0.54	0.64

Source: Own Compilation based on data from Thomson Reuters Data Stream (2016)

During the observed three-year period, the equal-weighted portfolio generates on average a return of -2.33% (annualized return). Compared to that, the value-weighted portfolio performs much better and generates an annualized return of 3.49%. Obviously, portfolio weights play an important role. Unfortunately, neither portfolio performs better than the market, which produces a return of 10.2% per year. *Figure 2* shows the accumulated underperformance of the two portfolios over the three-year period. The Sharpe Ratios[13] (SR) of the portfolios are also lower than the SR of the market (equal -1.72; value 0.54; market 0.64). The portfolio performance does not improve when risk-adjusted. The regression does not yield the desired results for the equal-weighted portfolio nor for the value-weighted one. The equal-weighted portfolio produces negative alphas in both cases: when regressed on the MRP, -4.2% p.a., and when regressed on all three risk factors, -4.1% p.a. The underperformance of the investment strategy is highly significant on the 1% level. For the value-weighted portfolio, the results differ between the two experimental designs. Jensen's Alpha is significantly negative (-0.56% p.a.), but Fama's Alpha is positive (0.46% p.a.). Unfortunately, the latter value is not

---

[13] Sharpe Ratio: performance measure which puts excess return in relation to the accompanied risk (Sharpe, 1964).

statistically significant. In other words, the experiment does not find evidence for a superior return by implementing a socially responsible investing strategy. The adjusted R^2 is in all cases below 40%, in the Fama-French case of course higher than in the CAPM case. Obviously, the return variation is not fully explained by the factors controlled for.

Figure 2: Accumulated Return of SRI-Portfolios vs. S&P 500
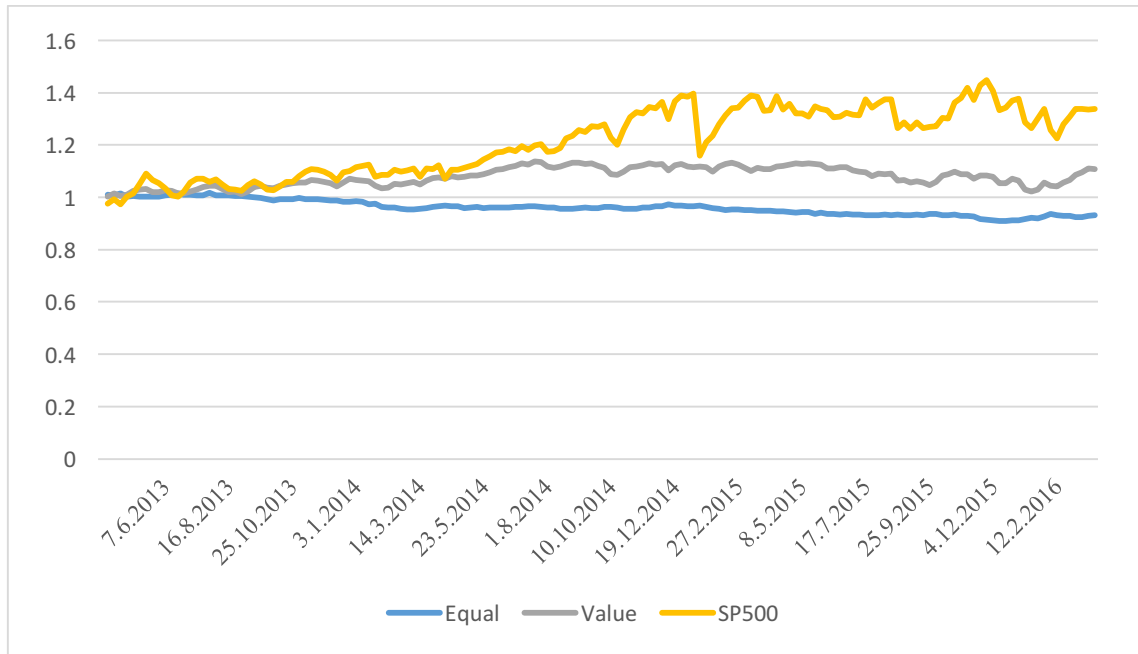
Source: Own Presentation based on data from Thomson Reuters Data Stream (2016)

Discussion and Limitations

Unfortunately, the empirical study does not produce the desired results. The return of the SRI strategy is not noticeably- or significantly superior to that of the market (S&P 500). The underperformance of the equal-weighted portfolio is probably affected by the weak performance of a large number of companies, which can be offset by the strong performance of big companies in the value-weighted portfolio.

These divergent results may have occurred for several reasons. The insignificance of the alpha in the value-weighted portfolio may result from an insufficient sample size. Another reason that this study could not find evidence for a positive alpha may be the time period observed, which is too short or unluckily chosen. The results might have been different if another region had been chosen to test the strategy. The ESG scores used here have to be addressed as well. The study used an overall score to capture sustainability as completely as possible. Other researchers have criticized such measures of CSP (Rowley & Berman, 2000). This argument

34

contradicts the findings of several authors (Derwall et al., 2005; Kempf & Osthoff, 2007), who based their research on KLD data (other ESG data provider), but it offers a reasonable explanation of the findings in this study. Regarding the two last-cited studies, the most important reason is probably a decisive difference in composition between the data from Sustainalytics and from KLD. Data from Sustainalytics were used here for their novelty and availability. The information advantage assumed to exist due to their novelty was not noticeable. R^2 is relatively small for all regressions, which is not unusual in such regression analyses. Nevertheless, this could be an indicator that other factors, such as R&D or other intangibles, may explain the real CSP-CFP linkage, as suggested by some researchers (McWilliams & Siegel, 2000). All in all, there could be many reasons for this outcome, and it is worth bearing in mind that the result of this study does not conclusively prove that SRI does *not* lead to a higher financial return due to several limitations.

The following points limit the expressiveness of the results and indicate caution for implications. First, the study was conducted as cross-sectional research that observed the performance of two portfolios over a three-year period. The period was chosen due to data restrictions but is too short for final long-term conclusions. Besides that, past related findings do not necessarily represent future stock scenarios. Second, the research focused on one specific asset class, stocks, in one specific area, USA. Accordingly, the results cannot be used as a generalization of SRI performance among all asset classes in the whole world. Third, the same overall score was used for the whole three-year period, again due to poor data availability. Fourth, this research did not directly address the causality of CSP on CFP or vice versa. What is more, CSP, represented by an overall score, was observed together with CFP, represented by stock returns, in a concurrent time frame, which did not allow for clear causal deductions. Lastly, this research did not consider dividends as part of shareholders return, nor transaction costs, which can significantly affect the total performance.

These limitations give reason to address these shortcomings in future research. The alleged superiority of KLD data is only hypothetical, but it would be interesting if an ensuing study tested the difference of the two CSP measures. Apart from that, future research should examine the effect of dividends as part of shareholders' return. Another topic which should be addressed by upcoming studies is the transaction costs which burden investors in a real environment.

Chapter 4 Conclusion

The purpose of this chapter is to highlight the most important findings of this study. The focus is thereby set on the literature review and the empirical study.

Environmental, social, and governance issues have gained great importance in recent years. Although the volume of SRI has increased steadily in the last decade, there is still disagreement about the financial return arising from these SRI investments among academic researchers. A variety of empirical findings of the implicated CSP-CFP relationship exists, and each of these findings can be explained by another theoretical framework. The most important are the social impact theory, which suggests a positive causal impact of CSP on CFP, and the trade-off theory, which claims a negative causal impact of CSP on CFP. The empirical findings are as various as the used methods and measures of primary and secondary studies are. With respect to this variety, secondary studies that review existing primary studies produce more meaningful results.

The execution of the literature review generated some interesting insights: first of all, it has to be mentioned that the heterogeneity within the secondary literature in this field is very challenging and makes it almost impossible to make clear comparisons between papers. The approach used in this work of categorizing papers according to their employed method and main purpose made it possible to obtain four categories of similar studies: narrative, theoretical, vote-count and meta-analysis studies. Since the quality of a paper, here understood as the sophistication of the research design, and the expressiveness of the concluded result are correlated, it is reasonable to evaluate the examined studies based on qualitative aspects. The four categories allow for such comparisons within and across category borders. It turns out that not all papers, respectively their results, comprise the same information and feature the same level of trustworthiness. In general, narrative reviews give a good (historical) overview of the existing literature, but cannot be used to conclude any final relationship between CSP and CFP. Theoretical papers show interesting details of the CSP-CFP relationship, when they construct theoretical frameworks to explain empirical findings. Similar to narrative reviews, these studies do not display empirical evidence for a certain relationship. Vote-count studies use a simple and comprehensible but unreliable and potential misleading approach to proof the linkage. Hence, it cannot be concluded a certain CSP-CFP relationship from this studies. A much more reliable methodology is applied by meta-analyses, which are the state-of-the-art studies at the moment. They consider and shed light on

all the qualitative important aspects of an elaborate study. Besides the methodological superiority of meta-analyses, the extensive data encompassed are another convincing argument for the trustworthiness of meta-analysis results. The fact that all reviewed meta-analyses, which comprise a huge amount of single results together, found a positive CSP-CFP relationship, give confidence that CSP indeed positively influences CFP.

The content of the reviewed papers gave evidence for the opinion of Margolis and Walsh (2003) cited above (p. 13): repeating criticism of successive studies without hope for relief and a conclusive result. But this is only the half the truth. This study suggests an alternative categorization of distinct studies based on qualitative characteristics which are related to the quality of the studies' results. The qualitative evaluation of these studies shows that we know more than other authors have claimed, but that we have focused on the wrong kind of studies so far. Sophisticated studies give evidence for a final positive CSP-CFP relationship.

The empirical study which addressed the research question, whether it is possible to achieve a superior return by pursuing a SRI strategy did not yield the desired results. The study aimed at finding further evidence for a risk-adjusted superior return of SRI. Instead, the results are divergent. The equal-weighted portfolio performed worse than the value-weighted portfolio and generated a significant under-return. Obviously, portfolio weights play an important role and a value-weighted portfolio should be chosen for SRI. Since the study only found insignificant evidence for a superior return of the value-weighted portfolio, no such excess return can be concluded with certainty. But this does also not prove the opposite. It is possible that SRI does indeed not generate a superior return, but it is also feasible that this result was caused by one or more research parameters chosen. Restricted data availability was one of the most challenging points of the study. Excess returns which were found by other authors who had employed different ESG data, show the importance of the scores' composition. Satisfying data availability and the quality of the employed ESG scores are probably the most crucial preconditions for a successful SRI strategy. The difference in quality between different ESG scores has to be addressed by further research to improve the performance of SRI and to give investors reliable strategies.

Taken all together, sophisticated reviews show a clear positive relationship between corporate social behavior and financial performance, which is the prerequisite for successful SRI. Many primary studies found superior returns of SRI strategies. The in this paper conducted

empirical study unfortunately did not find a significant positive superior return but was limited in many directions. Under appropriate circumstances SRI can outperform conventional investments. But even if SRI goes together with an inferior compensation, the urgency of a sustainable development will let this investment class further grow.

Chapter 5 Bibliography

Abbott, W. F., & Monsen, R. J. (1979). On the measurement of corporate social responsibility: Self-reported disclosures as a method of measuring corporate social involvement. *Academy of Management Journal*, *22*, 501–515.

Aguinis, H., & Glavas, A. (2012). What we know and don't know about corporate social responsibility a review and research agenda. *Journal of Management*, *38*, 932–968.

Ahmed, S. U., & Uchida, S. (2009). Corporate Social Responsibility and Financial Performance Linkage: A Preliminary Study for the Conceptual Framework (SSRN Scholarly Paper No. ID 2342568). Retrieved from http://papers.ssrn.com/abstract=2342568, 02.04.2016.

Albertini, E. (2013). Does Environmental Management Improve Financial Performance? A Meta-Analytical Review. *Organization & Environment*, *26*, 431–457.

Aldag, R. J., & Bartol, K. M. (1978). Empirical studies of corporate social performance and policy: A survey of problems and results. *Research in Corporate Social Performance and Policy*, *1*, 165–199.

Alexander, G. J., & Buchholz, R. A. (1978). Corporate social responsibility and stock market performance. *Academy of Management Journal*, *21*, 479–486.

Allouche, J., & Laroche, P. (2005). A meta-analytical investigation of the relationship between corporate social and financial performance. *Revue de Gestion Des Ressources Humaines*, *57*, 18–40.

Ambec, S., & Lanoie, P. (2008). Does it pay to be green? A systematic overview. *The Academy of Management Perspectives*, *22*, 45–62.

Arlow, P., & Gannon, M. J. (1982). Social responsiveness, corporate structure, and economic performance. *Academy of Management Review*, *7*, 235–241.

Auer, B. R., & Schuhmacher, F. (2016). Do socially (ir)responsible investments pay? New evidence from international ESG data. *The Quarterly Review of Economics and Finance*, *59*, 51–62.

Aupperle, K. E., Carroll, A. B., & Hatfield, J. D. (1985). An empirical examination of the relationship between corporate social responsibility and profitability. *Academy of Management Journal*, *28*, 446–463.

Barnett, M. L., & Salomon, R. M. (2006). Beyond dichotomy: The curvilinear relationship between social responsibility and financial performance. *Strategic Management Journal*, *27*, 1101–1122.

Beurden, P. van, & Gössling, T. (2008). The Worth of Values – A Literature Review on the Relation Between Corporate Social and Financial Performance. *Journal of Business Ethics*, *82*, 407–424.

Boaventura, J. M. G., Silva, R. S. da, & Bandeira-de-Mello, R. (2012). Corporate Financial Performance and Corporate Social Performance: methodological development and the theoretical contribution of empirical studies. *Revista Contabilidade & Finanças*, *23*, 232–245.

Bodie, Z., Kane, A., & Marcus, A. (2014). *Investments* (global ed. (3rd of January 2014)). Columbus: McGraw-Hill Higher Education.

Bowman, E. H., & Haire, M. (1975). A strategic posture toward corporate social responsibility. *California Management Review*, *18*, 49–58.

Bragdon Jr, J. H., & Marlin, J. (1972). ls Pollution Profitable? *Risk Management*, *19*, 9–18.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*, 57–82.

Carroll, A. B. (1979). A three-dimensional conceptual model of corporate performance. *Academy of Management Review*, *4*, 497–505.

Carroll, A. B., & Shabana, K. M. (2010). The business case for corporate social responsibility: A review of concepts, research and practice. *International Journal of Management Reviews*, *12*, 85–105.

Choi, J., & Wang, H. (2009). Stakeholder relations and the persistence of corporate financial performance. *Strategic Management Journal*, *30*, 895–907.

Clark, G. L., Feiner, A., & Viehs, M. (2015). From the Stockholder to the Stakeholder: How Sustainability Can Drive Financial Outperformance (SSRN Scholarly Paper No. ID 2508281). Retrieved from http://papers.ssrn.com/abstract=2508281, 02.04.16.

Clark, G. L., & Viehs, M. (2014). The Implications of Corporate Social Responsibility for Investors: An Overview and Evaluation of the Existing CSR Literature (SSRN Scholarly Paper No. ID 2481877). Retrieved from http://papers.ssrn.com/abstract=2481877, 05.04.2016.

Cochran, P. L., & Wood, R. A. (1984). Corporate social responsibility and financial performance. *Academy of Management Journal*, *27*, 42–56.

Dam, L. (2008). Corporate social responsibility and financial markets. University Library of Groningen.

Derwall, J., Guenster, N., Bauer, R., & Koedijk, K. (2005). The eco-efficiency premium puzzle. *Financial Analysts Journal*, *61*, 51–63.

Derwall, J., Koedijk, K., & Ter Horst, J. (2011). A tale of values-driven and profit-seeking social investors. *Journal of Banking & Finance*, *35*, 2137–2147.

Dictionary.com. (2016). *correlation*. Retrieved from http://www.dictionary.com/browse/correlation, 08.06.2016.

Endrikat, J., Guenther, E., & Hoppe, H. (2014). Making sense of conflicting empirical findings: A meta-analytic review of the relationship between corporate environmental and financial performance. *European Management Journal, 32*, 735–751.

Eurosif. (2014). *European SRI Study*.

Eurosif. (2016). Our Network, GSIA. Retrieved from http://www.eurosif.org/our-network/gsia/, 17.06.2016.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*, 3–56.

FNG. (2015). *Marktbericht Nachhaltige Geldanlagen 2015*, May.

FNG. (2016). Über uns, Das Forum Nachhaltige Geldanlagen e. V. Retrieved from http://www.forum-ng.org/de/fng/ueber-uns.html, 24.04.2016.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, *5*, 210–233.

Friedman, M. (1962). *Capitalism and Freedom*. University of Chicago Press.

Friedman, M. (1970). The social responsibility of business is to increase its profits. *New York,* 122–124.

Fulton, M., Kahn, B. M., & Sharples, C. (2012). Sustainable investing: Establishing long-term value and performance. Deutsche Bank Climate Change Advisors.

Grewatsch, S., & Kleindienst, I. (2015). When Does It Pay to be Good? Moderators and Mediators in the Corporate Sustainability–Corporate Financial Performance Relationship: A Critical Review. *Journal of Business Ethics*, 1–34.

Griffin, J. J., & Mahon, J. F. (1997). The corporate social performance and corporate financial performance debate twenty-five years of incomparable research. *Business & Society*, *36*, 5–31.

GSIA. (2015). *Global Sustainable Investment Review 2014*, February.

Harrison, J. S., & Freeman, R. E. (1999). Stakeholders, social responsibility, and performance: Empirical evidence and theoretical perspectives. *Academy of Management Journal, 42*, 479–485.

Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis.

Psychological Bulletin, *88*, 359–369.

Hoepner, A. G. F., & McMillan, D. G. (2009). Research on "Responsible Investment": An Influential Literature Analysis Comprising a Rating, Characterisation, Categorisation and Investigation (SSRN Scholarly Paper No. ID 1454793). Retrieved from http://papers.ssrn.com/abstract=1454793, 15.04.2016.

Hoffmann, J., Scherhorn, G., & Busch, T. (2004). *Darmstadt Definition of Sustainable Investments*. Wuppertal Institut für Klima, Umwelt, Energie GmbH.

Huppé, G. A. (2011). Alpha's Tale: The Economic Value of CSR (SSRN Scholarly Paper No. ID 1969583). Retrieved from http://papers.ssrn.com/abstract=1969583, 14.04.2016.

Investopedia. (2016). Impact Investing vs. Venture Philanthropy. Retrieved from http://www.investopedia.com/articles/personal-finance/060915/impact-investing-vs-venture-philanthropy.asp, 24.04.2016.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of Finance*, *23*, 389–416.

Kempf, A., & Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management*, *13*, 908–922.

Kenneth R. French. (2016). Current Research Returns. Retrieved from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#HistBenchmarks, 20.05.2016.

Kurtz, L. (2005). Answers to four questions. *The Journal of Investing*, *14*, 125–140.

Lankoski, L. (2000). Determinants of environmental profit: An analysis of the firm-level relationship between environmental performance and economic performance. Helsinki University of Technology.

Latinovic, M., & Obradovic, T. (2013). The Performance of Socially Responsible Investments. *Entrepreneurial Business and Economics Review*, *1*, 29–40.

Le Maux, J., & Le Saout, E. (2004). The performance of sustainability indexes. *Finance India*, *18*, 737–750.

Lerner, L. D., & Fryxell, G. E. (1988). An empirical study of the predictors of corporate social performance: A multi-dimensional analysis. *Journal of Business Ethics*, 951–959.

Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*, 383–417.

Margolis, J. D., Elfenbein, H. A., & Walsh, J. P. (2007). Does it pay to be good? A meta-analysis and redirection of research on the relationship between corporate social and financial performance. University of Michigan.

Margolis, J. D., Elfenbein, H. A., & Walsh, J. P. (2009). Does it Pay to Be Good...And Does it Matter? A Meta-Analysis of the Relationship between Corporate Social and Financial Performance (SSRN Scholarly Paper No. ID 1866371). Retrieved from http://papers.ssrn.com/abstract=1866371, 02.04.2016.

Margolis, J. D., & Walsh, J. P. (2003). Misery loves companies: Rethinking social initiatives by business. *Administrative Science Quarterly*, *48*, 268–305.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*, 77–91.

McGuire, J. B., Sundgren, A., & Schneeweis, T. (1988). Corporate social responsibility and firm financial performance. *Academy of Management Journal*, *31*, 854–872.

McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and financial performance: correlation or misspecification? *Strategic Management Journal*, *21*, 603–609.

McWilliams, A., Siegel, D. S., & Wright, P. M. (2006). Corporate social responsibility: Strategic implications*. *Journal of Management Studies*, *43*, 1–18.

Mercer, M., Carpenter, G., & Wyman, O. (2009). Shedding light on responsible investment: Approaches, returns and impact*s*. Mercer.

Moskowitz, M. (1972). Choosing socially responsible stocks. *Business and Society Review*, *1*, 71–75.

MSCI. (2016). What We Offer, ESG Integration. Retrieved from https://www.msci.com/esg-integration, 28.04.2016.

Orlitzky, M. (2001). Does Firm Size Comfound the Relationship Between Corporate Social Performance and Firm Financial Performance? *Journal of Business Ethics*, *33*, 167–180.

Orlitzky, M., & Benjamin, J. D. (2001). Corporate social performance and firm risk: A meta-analytic review. *Business & Society*, *40*, 369–396.

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate Social and Financial Performance: A Meta-Analysis. *Organization Studies*, *24*, 403–441.

Pava, M. L., & Krausz, J. (1996). The association between corporate social-responsibility and financial performance: The paradox of social cost. *Journal of Business Ethics*, *15*, 321–357.

Peloza, J. (2009). The challenge of measuring financial impacts from investments in corporate social performance. *Journal of Management*, *35*, 1518–1541.

Perrini, F., Russo, A., Tencati, A., & Vurro, C. (2012). Deconstructing the Relationship Between Corporate Social and Financial Performance. *Journal of Business Ethics*, *102*, 59–76.

Posner, B. Z., & Schmidt, W. H. (1992). Values and the American manager: An update updated. *California Management Review*, *34*, 80–94.

Preston, L. E., & O'bannon, D. P. (1997). The corporate social-financial performance relationship. *Business and Society*, *36*, 419–429.

PRI. (2016). About the PRI, The Six Principles. Retrieved from https://www.unpri.org/about/the-six-principles, 24.04.2016.

Raar, J. (2001). Strategy and the multiplicity of variables associated with voluntary reporting of environmental and economic performance: Ullmann revisited. *Governance and Corporate Social Responsibility in the New Millennium : Governance and Social Responsibility Conference : Proceedings of the 2001 Conference, 26th & 27th November 2001*, 212–221.

Rathner, S. (2012). The Influence of Primary Study Characteristics on the Performance Differential Between Socially Responsible and Conventional Investment Funds: A Meta-Analysis. *Journal of Business Ethics*, *118*, 349–363.

Raza, A., Ilyas, M. I., Rauf, R., & Qamar, R. (2012). Relationship between Corporate Social Responsibility (CSR) and Corporate Financial Performance (CFP): Literature review approach. *Elixir Financ. Manag*, *46*, 8404–8409.

Renneboog, L., Ter Horst, J., & Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance*, *32*, 1723–1742.

Revelli, C., & Viviani, J. (2015). Financial performance of socially responsible investing (SRI): what have we learned? A meta-analysis. *Business Ethics: A European Review*, *24*, 158–185.

Revelli, C., & Viviani, J.-L. (2013). The link between SRI and financial performance: effects and moderators. *Management international/International Management/Gestiòn Internacional*, *17*, 105–122.

Roman, R. M., Hayibor, S., & Agle, B. R. (1999). The relationship between social and financial performance repainting a portrait. *Business & Society*, *38*, 109–125.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Vol. 6). London: Sage publications.

Rowley, T., & Berman, S. (2000). A brand new brand of corporate social performance. *Business & Society*, *39*, 397–418.

Salzmann, O., Ionescu-Somers, A., & Steger, U. (2005). The business case for corporate sustainability:: literature review and research options. *European Management Journal*,

23, 27–36.

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (Vol. 3). London: Sage publications.

Schreck, P. (2008). Erfolgswirkungen von Umwelt-und Sozialstandards, Empirische Erkenntnisse zu Corporate Social Responsibility als Business Case. *Zweiter Runder Tisch Bayern, Sozial-Und Umweltstandards Bei Unternehmen. München/Augsburg (Bay. Landeszentrale Für Politische Bildung)*, 87–101.

Schröder, M. (2014). Financial effects of corporate social responsibility: a literature review. *Journal of Sustainable Finance & Investment*, *4*, 337–350.

Shane, P. B., & Spicer, B. H. (1983). Market Response to Environmental Information Produced outside the Firm. *The Accounting Review*, *58*, 521–538.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*, 425–442.

Sjöström, E. (2011). The Performance of Socially Responsible Investment - A Review of Scholarly Studies Published 2008-2010 (SSRN Scholarly Paper No. ID 1948169). Retrieved from http://papers.ssrn.com/abstract=1948169, 14.04.2016.

Statistics Solutions. (2016). One Sample t-Test. Retrieved from http://www.statisticssolutions.com/manova-analysis-one-sample-t-test/, 06.05.2016.

Sustainalytics. (2016). What we do, Our approach. Retrieved from http://www.sustainalytics.com/our-approach, 28.04.2016.

Thomson Reuters. (2016). Data & Analytics, ESG Research Data. Retrieved from http://financial.thomsonreuters.com/en/products/data-analytics/company-data/esg-research-data.html, 28.04.2016.

Ullmann, A. A. (1985). Data in search of a theory: A critical examination of the relationships among social performance, social disclosure, and economic performance of US firms. *Academy of Management Review*, *10*, 540–557.

UNEP Finance Initiative. (2006). Show Me The Money: Linking Environmental, Social and Governance Issues to Company Value.

UNEP Finance Initiative. (2007). Demystifying responsible investment performance- A review of key academic and broker research on ESG factors.

Universität Zürich. (2016). Datenanalyse, Multiple Regressionsanalyse. Retrieved from http://www.methodenberatung.uzh.ch/datenanalyse/zusammenhaenge/mreg.html, 06.06.2016.

Vance, S. C. (1975). Are socially responsible corporations good investment risks.

Management Review, *64*, 18–24.

Waddock, S. A., & Graves, S. B. (1997). The corporate social performance-financial performance link. *Strategic Management Journal*, *18*, 303–319.

Wagner, M. (2001). A review of empirical studies concerning the relationship between environmental and economic performance: What does the evidence tell us? Centre for Sustainability Management.

Wallis, M. von, & Klein, C. (2014). Ethical requirement and financial interest: a literature review on socially responsible investing. *Business Research*, *8*, 61–98.

Wang, Q., Dou, J., & Jia, S. (2015). A Meta-Analytic Review of Corporate Social Responsibility and Corporate Financial Performance The Moderating Effect of Contextual Factors. *Business & Society*, 1–39.

Wood, D. J. (2010). Measuring corporate social performance: A review. *International Journal of Management Reviews*, *12*, 50–84.

World Business Council for Sustainable Development. (1999). Corporate Social Responsibility: Meeting Changing Expectations.

Database

Bloomberg. (2016). Request by means of Bloomberg Terminal, retrieved in April 2016.

Thomson Reuters Data Stream. (2016). Request by means of Data Stream Portal, retrieved in April 2016.

Chapter 6 Appendix

6.1 Excluded Studies

Table 2: Excluded Studies

E studies	S studies
Albertini (2013)	Arlow & Gannon (1982)
Ambec & Lanoie (2008)	Combs, Liu, Hall, & Ketchen (2006)
Blanco, Rey-Maquieira, & Lozano (2009)	Crook, Todd, Combs, Woehr, & Ketchen Jr (2011)
Dixon-Fowler, Slater, Johnson, Ellstrand, & Romi (2012)	Mayer-Haug, Read, Brinckmann, Dew, & Grichnik (2013)
Endrikat (2015)	Stam, Arzlanian, & Elfring (2014)
Endrikat et al. (2014)	Unger, Rauch, Frese, & Rosenbusch (2011)
Golicic & Smith (2013)	Van Wijk, Jansen, & Lyles (2008)
Guenther, Hoppe, & Endrikat (2012)	Westlund & Adam (2010)
Horváthová (2010)	
Lankoski (2000)	
Schultze & Trommer (2012)	
Shane & Spicer (1983)	
Wagner (2001)	
White & Matthew (2004)	

G studies
Gillan, Hartzell, Koch, & Starks (2011)
Love (2010)

Source: own compilation

6.2 Literature Review

6.2.1 Narrative Reviews

Table 3: Narrative Reviews

Study	Selection Criteria	Method	Considered period	Causality	Endogeneity	Comparability	Content/Procedure Comments
Aldag & Bartol (1978)	not disclosed	explained	1972-1978	considered	not considered	Study has a general view and does not specifically distinguish between papers	Study gives an incomplete overview of existing papers paired with recommendations for further research
Cochran & Wood (1984)	not disclosed	explained	1972-1982	not considered	considered and implemented in own study	Study does not systematically categorize papers to draw conclusions from homogeneous findings	Study uses review to justify own study (interesting result: asset-age and asset-turnover as significant control variables but positive CSP-CFP relationship still remained significant positive)
Aupperle et al. (1985)	not disclosed	explained	1972-1979	not considered	considered	Study examines methodologies and performance measures, but does not systematically categorize papers to draw conclusions from homogeneous findings	Study uses review to justify own study
Raar (2001)	disclosed	explained	1970-2000	not considered	considered	Study categorizes papers according to Ullmann: -> social disclosure and social performance -> social performance and economic performance -> social disclosure and economic performance	Results rather unclear and pointless
UNEP (2006)	not disclosed	explained	not disclosed	not considered	not considered	Study does not distinguish between papers	Narrative and tabularly summary of case studies Conclusion unreliable and based on single case studies Seems to be commercial-driven
UNEP (2007)	disclosed	explained	1996-2000	not considered	not considered	Study categorizes academic studies according to E, S and G dimensions	Narrative and tabularly summary of 20 academic studies and 10 broker (case) studies Inconclusive results from the academic papers and inexpressive conclusion of broker studies Seems to be commercial-driven
Renneboog et al. (2008)	not disclosed	explained	1992-2007	implemented	not considered	Fund-study categorizes according to regions	Study summarizes literature on firm and portfolio level Study separately summarizes primary studies which explicitly examine causality Study checks for regional differences Study gives a deliberate overview of SRI
Ahmed & Uchida (2009)	not disclosed	explained	1972-2007	considered	implemented	Study does not properly categorize primary studies	Study structure and methodology is not comprehensible Conclusion is pointless

Study	Selection Criteria	Method	Considered period	Causality	Endogeneity	Comparability	Content/Procedure Comments
Hoepner & McMillan (2009)	disclosed	explained	1991-2007	not considered	considered	Selection based on new influential literature analysis (ILA) (literature selection process based on citations)	Study draws conclusions from 51 influential papers Procedure could produce interesting results, since influence should be an indicator for reliability, but could also be suspect, because these conclusions had been drawn from a small number of single studies Study examines methodology, sample-sizes, control variables, significance and criticizes certain papers Study is conducted carefully
Mercer et al. (2009)	disclosed	explained	2006-2009	not considered	not considered	Study discriminates studies based on E, S and G dimension	Short summary of research, which follows up prior research UNEP (2007) Study presents tabularly results and deduces implications differentiated according to E, S and G dimension Seems to be commercial-driven
Derwall et al. (2011)	not disclosed	explained	1971-2010	not considered	considered	Fund-study: study categorizes studies based on the implemented screening approach (negative-screening vs. others)	Study proves "error-in-expectation"- and "shun-stock"-hypothesis by means of existing literature and own study with interesting results and implications for investors Conclusion makes sense
Huppé (2011)	not disclosed	explained	1972-2010	considered and implemented in own study	considered and implemented in own study	Study categorizes papers in accounting and market-based CFP studies and according to E and S dimension and overall scores	Study gives narrative overview arranged according to E, S and overall CSP performance Study conclusion not comprehensible (lack of fundament) Study addresses CSP-CFP in own regression analysis
Latinovic & Obradovic (2013)	not disclosed	explained	2008-2014	not considered	not considered	Fund-Study, no specific distinction employed	Summary of 8 single regional distinct fund studies, which are not analyzed for risk-adjusted returns etc. Conclusion is inexpressive
Clark & Viehs (2014)	not disclosed	explained	1988-2012	considered	considered	Study categorizes papers in accounting and market-based CFP studies and according to E, S and G dimension and overall scores	Study explains intention and procedure in an understandable way Study draws conclusions based on homogeneous study sample Study examines regional and industry-specific differences
Schröder (2014)	not disclosed	explained	2005-2014	considered	considered	Fund-study distinguishes between SRI-portfolio studies, SRI-indices and long-short strategy portfolios	Conclusion is not comprehensible

Causality: *not considered* if Author does not show awareness, *considered* if Author shows awareness, *implemented* if Author explicitly investigates causality, *implemented in own study* if Author investigates causality in own empirical study; Endogeneity: *not considered* if Author does not show awareness, *considered* if Author shows awareness (e.g. consideration of control variables, moderators, mediators etc.), *implemented* if Author explicitly investigates moderating/ mediating effects, *implemented in own study* if Author investigates moderating/ mediating effects in own empirical study; Selection Criteria: *disclosed* if Author presents selection criteria/process, *not disclose*d if Author does not present selection criteria/process; Method: *explained* if Author explains procedure and purpose, *not explained* if Author does not explicitly mention study procedure

Source: Own Compilation

6.2.2 Theoretical Reviews

Table 4: Theoretical Reviews

Study	Selection Criteria	Method	Data Period	Causality	Endogeneity	Comparability	Content/Procedure Comments
Harrison & Freeman (1999)	not disclosed	explained	1970-1998	not considered	not considered	-	Study focuses on only 6 papers and comments on event-studies, case methods, data base and data search
Rowley & Berman (2000)	not disclosed	explained	1962-1998	implemented	not considered	Study categorizes papers in single- and multi-dimension CSP measure papers	Study criticizes CSP construct and suggests new research direction, which should investigate when which CSP-CFP relationship can be expected and under what conditions Study examines shareholder activism as an important example and driver of the causal CSP-CFP linkage Authors raises interesting aspects for future research
Salzmann et al. (2005)	not disclosed	explained	1972-2004	considered	not considered	Study categorizes studies in empirical and theoretical literature	Study gives an overview over currently available frameworks, descriptive and empirical studies, as well as tools
Peloza (2009)	disclosed	explained	1975-2008	implemented (CSP->CFP)	considered	Study categorizes papers according to CSP/CFP measures	Vote-count result calculated from heterogeneous study sample Study examines employed CSP and CFP measures on different levels (mediating metrics, intermediate outcome metrics and end state outcome metrics) to deduce a theoretical causal CSP->CFP framework
Perrini et al. (2012)	not disclosed	explained	1970-2012	implemented (CSP->CFP)	considered	Study categorizes papers according to topic related stakeholder groups	Study analyzes CSR/CSP-CFP literature for key-drivers to deduce a theoretical (causal) framework from CSP to CFP
Rathner (2012)	disclosed	explained	1981-2008	implemented	implemented	Fund-study does not distinguish between papers for vote-count calculation Fund-study distinguishes between under- and outperformance studies for regression analysis	Study uses a multivariate regression approach to explain prior mixed study results of SRI-fund studies -> study characteristics as explaining variables Study reveals comprehensible results
Grewatsch & Kleindienst (2015)	disclosed	explained	1972-2014	implemented (CSP->CFP)	implemented	Study categorizes studies based on CSP measure, moderators, mediators and CFP measure	Study focuses on unilateral causal relationship (CSP->CFP) and addresses endogeneity aspects Result is a comprehensible theoretical framework deduced from empirical findings

Causality: *not considered* if Author does not show awareness, *considered* if Author shows awareness, *implemented* if Author explicitly investigates causality, *implemented in own study* if Author investigates causality in own empirical study; Endogeneity: *not considered* if Author does not show awareness, *considered* if Author shows awareness (e.g. consideration of control variables, moderators, mediators etc.), *implemented* if Author explicitly investigates moderating/ mediating effects, *implemented in own study* if Author investigates moderating/ mediating effects in own empirical study; Selection Criteria: *disclosed* if Author presents selection criteria/process, *not disclose*d if Author does not present selection criteria/process; Method: *explained* if Author explains procedure and purpose, *not explained* if Author does not explicitly mention study procedure
Source: Own Compilation

6.2.3 Vote-Count Reviews

Table 5: Vote-Count Reviews

Study	Selection Criteria	Method	Data Period	Causality	Endogeneity	Comparability	Content/Procedure Comments
Ullmann (1985)	not disclosed	explained	1972-1985	considered	considered	Study categorizes papers according to examined relationship: -> social disclosure and social performance -> social performance and economic performance -> social disclosure and economic performance	One of the most impactful studies in this field and has been cited by almost every succeeding study Result based on heterogeneous studies (different CSP measures etc.) Study creates three-dimensional model to explain inconclusive results. According to the author the missing element had been strategy until then. Hence, that was the element which had to be embedded in the theory
Pava & Krausz (1996)	not disclosed	explained for own study, not for review	1972-1992	not considered	not considered	Study categorizes papers according to CSP/CFP measures	Result based on heterogeneous (and homogeneous) studies Study combines distinct primary studies on purpose Study addresses sample-size problem Vote-count analysis based on 21 randomly matched studies is not comprehensible
Griffin & Mahon (1997)	disclosed	explained	1972-1995	not considered	considered	Study does not distinguish between papers	Study analyzes papers for employed CSP/CFP measures, control variables and significance Wrong indication of certain correlations -> biased vote-count result Result based on heterogeneous studies
Roman et al. (1999)	disclosed	explained	1972-1999	not considered	not considered	Study does not distinguish between papers	Study corrects results of Griffin & Mahon (1997) and reclassifies some of the papers Vote-count result based on heterogeneous studies The reclassification/exclusion of certain papers is object of subjectivity and does not improve the reliability of the final outcome
Margolis & Walsh (2003)	not disclosed	not explained	1972-2002	considered	not considered	Study discriminates studies based on CSP/CFP measures and CSP as dependent/independent variable	Result based on heterogeneous studies and pointless The unreliable result is then used as starting point of the discussion about organizational theory
Beurden & Gössling (2008)	disclosed	explained	1990-2006	not considered	implemented	Study categorizes studies according CSP/CFP measures	Study considers only primary studies after 1990 Study analyzes moderators Study draws conclusion from heterogeneous study sample
Dam (2008)	disclosed	explained	1976-2007	considered	not considered	Study categorizes studies according CFP measure	Study establishes own equilibrium model to explain precedent found vote-count results
Sjödström (2011)	disclosed	explained	2008-2010	not considered	not considered	Unclear whether fund-study screens only for risk-adjusted studies	Result possibly based on heterogeneous studies Author does not deal with studies in detail Conclusion inexpressive

Study	Selection Criteria	Method	Data Period	Causality	Endogeneity	Comparability	Content/Procedure Comments
Boaventura et al. (2012)	disclosed	explained	1996-2010	considered	considered	Study does not discriminate between studies	Study considers only primary studies after 1996 Study gives descriptive overview of used methods and measures Study draws conclusion from heterogeneous study sample
Fulton et al. (2012)	disclosed	explained	1991-2011	considered	considered	Study discriminates studies based on CSR/SRI/ESG character, and E, S and G dimension	Study explains procedure but gave an unorganized impression Results based on homogeneous studies (but often on a small number of studies) Study seems to be commercial-driven
Raza et al. (2012)	not disclosed	not explained	1972-2012	not considered	not considered	Study categorizes studies according to CFP measure	Categorized studies are blindly counted without analyzing underlying studies Result based on homogeneous studies Conclusion is unreliable and pointless
Revelli & Viviani (2013)	disclosed	explained	1972-2009	not considered	implemented	Fund-study categorizes studies in published/unpublished papers	Vote-Count result based on heterogeneous studies (e.g. not only on studies with risk-adjusted returns) Study checks for publication bias Study analyzes moderators by means of OLS-regression Study reveals interesting fact, that SRI portfolios constructed by academic researcher often show superior returns -> hidden intentions
Wallis & Klein (2014)	disclosed	not explained	1986-2012	not considered	considered	Fund-study does not distinguish between papers	Study examines CSP-CFP on firm- and on portfolio-level Study examines employed CFP measures and sample sizes but calculates vote-count result from a undifferentiated study sample-> incomprehensible conclusion Study considers regional and time aspects of SRI-fund studies
Clark et al. (2015)	not disclosed	not explained	1972-2014	not considered	not considered	Study categorizes studies according E, S and G dimension and CFP measure	Result based on homogeneous study sample Seems to be commercial-driven

Causality: *not considered* if Author does not show awareness, *considered* if Author shows awareness, *implemented* if Author explicitly investigates causality, *implemented in own study* if Author investigates causality in own empirical study; Endogeneity: *not considered* if Author does not show awareness, *considered* if Author shows awareness (e.g. consideration of control variables, moderators, mediators etc.), *implemented* if Author explicitly investigates moderating/ mediating effects, *implemented in own study* if Author investigates moderating/ mediating effects in own empirical study; Selection Criteria: *disclosed* if Author presents selection criteria/process, *not disclose*d if Author does not present selection criteria/process; Method: *explained* if Author explains procedure and purpose, *not explained* if Author does not explicitly mention study procedure

Source: Own Compilation

6.2.4 Meta-Analyses

Table 6: Meta-Analyses

Study	Selection Criteria	Method	Data Period	Causality	Endogeneity	Comparability	Content/Procedure Comments
Orlitzky (2001)	disclosed	explained	1970-2000	implemented	implemented	Study discriminates studies based on size/CSP and CFP measures	41 underlying primary studies Study controls for size. CSP-CFP relation remains significant positive Statistical methods: Meta-analysis (Hunter & Schmidt) File drawer analysis Study documents research well
Orlitzky & Benjamin (2001)	disclosed	explained	1969-2000	implemented	not considered	Study discriminates studies based on CSP/financial risk measures	18 underlying primary studies Statistical methods: Meta-analysis (Hunter & Schmidt) File drawer analysis Study documents research well
Orlitzky et al. (2003)	disclosed	explained	1970-2000	considered	considered	Study discriminates studies based on CSP/CFP measures	52 underlying primary studies Study examines separately CSP-CFP without E and survey-/reputation measures Statistical methods: Meta-analysis (Hunter & Schmidt) File drawer analysis Study documents research very well
Allouche & Laroche (2005)	disclosed	explained	1972-2003	implemented	implemented	Study discriminates studies based on CSP/CFP measures	Study expands Orlitzky et al. (2003) approach with non-US studies and additional tests 82 underlying primary studies Statistical methods: Meta-analysis (Hunter & Schmidt) Meta-regression approach (to verify moderators) Meta significance testing (to avoid publication bias) Study documents research well
Margolis et al. (2007)	disclosed	explained	1972-2007	implemented	considered	Study discriminates studies based on CSP/CFP measures	167 underlying primary studies Study distinguishes between 9 CSP measures and 2 CFP measures Study examines implemented control variables Statistical methods: Meta-analysis (Hunter & Schmidt) File drawer analysis

Study	Selection Criteria	Method	Data Period	Causality	Endogeneity	Comparability	Content/Procedure Comments
Margolis et al. (2009)	disclosed	explained	1972-2007	implemented	implemented	Study discriminates studies based on CSP/CFP measures	214 underlying primary studies Study distinguishes between 9 CSP measures and 2 CFP measures Study examines implemented control variables (result: no effect!) Study conducts additional study with papers after 1998 Statistical methods: Meta-analysis (Hunter & Schmidt) File drawer analysis
Friede et al. (2015)	disclosed	explained	1972-2014	not considered	considered	Study distinguishes between vote-count and meta-analysis, regional differences, asset-classes and portfolio vs. non-portfolio studies and E, S and G dimension	Third level study Data from more than 2000 primary studies Study produced probably the most comprehensive results in this field so far. Study explains procedures in comprehensible way Study examines the CSP-CFP link under various different conditions (region, asset-class, portfolio vs. non-portfolio, E, S and G dimension) Statistical methods: Meta-analysis (Hunter & Schmidt) Binomial tests Effect sizes of vote count studies
Revelli & Viviani (2015)	disclosed	explained	1972-2012	not considered	implemented	Fund-study does not distinguish between studies	85 underlying primary studies Study objects the general assumption of underperforming SRI-funds vs. Conventional funds Study incorporates studies from the whole world Statistical methods: Meta-analysis (Hunter & Schmidt) Funnel plot analysis (to avoid publication bias) Procedure is well explained Conclusion is reasonable

Study	Selection Criteria	Method	Data Period	Causality	Endogeneity	Comparability	Content/Procedure Comments
Wang et al. (2015)	disclosed	explained	2003-2012	implemented	implemented	Study discriminates studies based on CSP/CFP measures	42 underlying primary studies Study calculates correlation for 5 CSP measure categories and 3 CFP measure categories, as well as for developed and undeveloped countries Study finds only evidence for unilateral causal relationship (CSP->CFP) Statistical methods: Meta-analysis (Hunter & Schmidt) File drawer analysis (to avoid publication bias) Homogeneity test Procedure is elaborate and well explained Conclusion is reasonable

Causality: *not considered* if Author does not show awareness, *considered* if Author shows awareness, *implemented* if Author explicitly investigates causality, *implemented in own study* if Author investigates causality in own empirical study; Endogeneity: *not considered* if Author does not show awareness, *considered* if Author shows awareness (e.g. consideration of control variables, moderators, mediators etc.), *implemented* if Author explicitly investigates moderating/ mediating effects, *implemented in own study* if Author investigates moderating/ mediating effects in own empirical study; Selection Criteria: *disclosed* if Author presents selection criteria/process, *not disclose*d if Author does not present selection criteria/process; Method: *explained* if Author explains procedure and purpose, *not explained* if Author does not explicitly mention study procedure
Source: Own Compilation

6.3 Empirical Study

6.3.1 Portfolio Structure

Table 7: Portfolio Structure

Ticker	Short Name	Ticker	Short Name
Long positions		**Short positions**	
Basic Materials			
PX	PRAXAIR INC	NUE	NUCOR CORP
APD	AIR PRODS & CHEM	LYB	LYONDELLBASELL-A
MOS	MOSAIC CO/THE	DD	DU PONT (EI)
FMC	FMC CORP	ARG	AIRGAS INC
ECL	ECOLAB INC	CF	CF INDUSTRIES HO
Consumer Goods			
CPB	CAMPBELL SOUP CO	LUK	LEUCADIA NATL
JCI	JOHNSON CONTROLS	LEG	LEGGETT & PLATT
K	KELLOGG CO	LEN	LENNAR CORP-A
MDLZ	MONDELEZ INTER-A	DHI	DR HORTON INC
PEP	PEPSICO INC	MNST	MONSTER BEVERAGE
Consumer Services			
BBY	BEST BUY CO INC	COST	COSTCO WHOLESALE
SBUX	STARBUCKS CORP	WYNN	WYNN RESORTS LTD
SPLS	STAPLES INC	NFLX	NETFLIX INC
RCL	ROYAL CARIBBEAN	EXPE	EXPEDIA INC
JWN	NORDSTROM INC	CVC	CABLEVISION SY-A
Financials			
NTRS	NORTHERN TRUST	L	LOEWS CORP
MHFI	MCGRAW HILL FINA	EFX	EQUIFAX INC
STT	STATE ST CORP	AMG	AFFIL MANAGERS
PLD	PROLOGIS INC	EXR	EXTRA SPACE STOR
GS	GOLDMAN SACHS GP	TMK	TORCHMARK CORP
Health Care			
JNJ	JOHNSON&JOHNSON	ALXN	ALEXION PHARM
BAX	BAXTER INTL INC	LH	LABORATORY CP
VAR	VARIAN MEDICAL S	ISRG	INTUITIVE SURGIC
MRK	MERCK & CO	UHS	UNIVERSAL HLTH-B
UNH	UNITEDHEALTH GRP	ENDP	ENDO INTERNATION
Industrials			
GWW	WW GRAINGER INC	FLIR	FLIR SYSTEMS
MMM	3M CO	ROP	ROPER TECHNOLOGI
ACN	ACCENTURE PLC-A	JBHT	HUNT (JB) TRANS
XRX	XEROX CORP	VRSK	VERISK ANALYTI
UPS	UNITED PARCEL-B	AME	AMETEK INC
Oil&Gas			
SE	SPECTRA ENERG	RIG	TRANSOCEAN LTD
BHI	BAKER HUGHES INC	XEC	CIMAREX ENERGY C
OXY	OCCIDENTAL PETE	HP	HELMERICH & PAYN
HES	HESS CORP	CXO	CONCHO RESOURCES
SLB	SCHLUMBERGER LTD	PSX	PHILLIPS 66
Technology			
EMC	EMC CORP/MA	LLTC	LINEAR TECH CORP
INTC	INTEL CORP	GRMN	GARMIN LTD
ADBE	ADOBE SYS INC	HRS	HARRIS CORP
SYMC	SYMANTEC CORP	QRVO	QORVO INC
IBM	IBM	FFIV	F5 NETWORKS
Telecommunications			
VZ	VERIZON COMMUNIC	FTR	FRONTIER COMMUNI
T	AT&T INC	LVLT	LEVEL 3 COMM INC
Utilities			
EXC	EXELON CORP	SCG	SCANA CORP
AWK	AMERICAN WATER W	CNP	CENTERPOINT ENER
CMS	CMS ENERGY CORP	PPL	PPL CORP
PCG	PG&E CORP	GAS	AGL RESOURCES
OKE	ONEOK INC	FE	FIRSTENERGY CORP

Source: Own Compilation based on data from Bloomberg (2016)

6.3.2 Regression Outputs

Table 8: Regression Output Equal-Weighted (CAPM)

Min	1Q	Median	3Q	Max
-0.0118676	-0.0020385	0.0001585	0.0023184	0.0094147

Coefficients:

	Estimate	Std. Error	t value Pr(>|t|)
Alpha	-0.0008041	0.0002822	-2.849 0.00498 **
MRP	-0.0464320	0.0104710	-4.434 1.75e-05 ***

Residual standard error:	0.003517 on 154 degrees of freedom		
Multiple R-squared:	0.1132,	Adjusted R-squared: 0.1075	
F-statistic:	19.66 on 1 and 154 DF,	p-value:	1.748e-05

heteroscedastic-consistent Std. Errors

	Estimate	Std. Error	t value	Pr(>|t|)
Alpha	-0.00080412	0.00028461	-2.8254	0.0053489 **
MRP	-0.04643199	0.01263057	-3.6762	0.0003265 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Source: Own Compilation based on data from Thomson Reuters Data Stream (2016)

Table 9: Regression Output Equal-Weighted (Fama-French)

Min	1Q	Median	3Q	Max
-0.0125110	-0.0018567	0.0000782	0.0021736	0.0093636

	Estimate	Std. Error	t value Pr(>|t|)
Alpha	-0.0008157	0.0002735	-2.982 0.00333 **
MRP	-0.0432567	0.0101375	-4.267 3.47e-05 ***
SMB	-0.0679178	0.0266654	-2.547 0.01186 *
HML	0.0587658	0.0286578	2.051 0.04202 *

Residual standard error:	0.003392 on 152 degrees of freedom		
Multiple R-squared:	0.1859,	Adjusted R-squared: 0.1698	
F-statistic:	11.57 on 3 and 152 DF,	p-value:	7.145e-07

heteroscedastic-consistent Std. Errors

	Estimate	Std. Error	t value	Pr(>|t|)
Alpha	-0.00081568	0.00027159	-3.0034	0.0031233 **
MRP	-0.04325670	0.01154616	-3.7464	0.0002542 ***
SMB	-0.06791778	0.02336941	-2.9063	0.0042048 **
HML	0.05876577	0.03387907	1.7346	0.0848434 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Source: Own Compilation based on data from Thomson Reuters Data Stream (2016)

Table 10: Regression Output Value-Weighted Portfolio (CAPM)

Min	1Q	Median	3Q	Max
-0.021535	-0.004793	0.000150	0.004757	0.036357

	Estimate	Std. Error	t value Pr(>|t|)
Alpha	-0.0001072	0.0006343	-0.169 0.866
MRP	0.2068451	0.0235329	8.790 2.78e-15 ***

Residual standard error: 0.007904 on 154 degrees of freedom
Multiple R-squared: 0.3341,   Adjusted R-squared: 0.3298
F-statistic: 77.26 on 1 and 154 DF   p-value: 2.779e-15

heteroscedastic-consistent Std. Errors

	Estimate	Std. Error	t value	Pr(>|t|)
Alpha	-0.00010719	0.00067894	-0.1579	0.8747600
MRP	0.20684511	0.06011172	3.4410	0.0007461 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Source: Own Compilation based on data from Thomson Reuters Data Stream (2016)

Table 11: Regression Output Value-Weighted Portfolio (Fama-French)

Min	1Q	Median	3Q	Max
-0.021112	-0.004981	0.000014	0.004486	0.036497

	Estimate	Std. Error	t value Pr(>|t|)
Alpha	8.738e-05	6.134e-04	0.142 0.886920
MRP	2.083e-01	2.274e-02	9.161 3.3e-16 ***
SMB	8.384e-02	5.980e-02	1.402 0.162994
HML	2.394e-01	6.427e-02	3.725 0.000275 ***

Residual standard error: 0.007607 on 152 degrees of freedom
Multiple R-squared: 0.3912,   Adjusted R-squared: 0.3792
F-statistic: 32.55 on 3 and 152 DF,   p-value: 2.612e-16

heteroscedastic-consistent Std. Errors

	Estimate	Std. Error	t value	Pr(>|t|)
Alpha	8.7376e-05	6.7501e-04	0.1294	0.8971784
MRP	2.0829e-01	6.0966e-02	3.4166	0.0008136 ***
SMB	8.3837e-02	5.8546e-02	1.4320	0.1541964
HML	2.3942e-01	7.0013e-02	3.4197	0.0008051 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Source: Own Compilation based on data from Thomson Reuters Data Stream (2016)