FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO

BARBARA WOEGERBAUER

COMPENSATION EFFECTS IN A SOCIALLY RESPONSIBLE INVESTMENT CONTEXT

SÃO PAULO
2019

BARBARA WOEGERBAUER

COMPENSATION EFFECTS IN A SOCIALLY RESPONSIBLE INVESTMENT CONTEXT

Thesis presented to Escola de Administração de Empresas de São Paulo of Fundação Getulio Vargas, as a requirement to obtain the title of Master in International Management (MPGI).

Knowledge Field: Economia e Finanças Internacionais

Adviser: Prof. Dr. Rafael Schiozer

SÃO PAULO
2019

Woegerbauer, Barbara.
Compensation effects in a Socially Responsible Investment context / Barbara Woegerbauer. - 2019.
68 f.

Orientador: Rafael Felipe Schiozer.
Dissertação (mestrado profissional MPGI) – Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo.

1. Investimentos - Aspectos sociais. 2. Investimentos - Aspectos ambientais. 3. Investimentos - Análise. 4. Responsabilidade social da empresa. I. Schiozer, Rafael Felipe. II. Dissertação (mestrado profissional MPGI) – Escola de Administração de Empresas de São Paulo. III. Fundação Getulio Vargas. IV. Título.

CDU 336.767

Ficha Catalográfica elaborada por: Isabele Oliveira dos Santos Garcia CRB SP-010191/O
Biblioteca Karl A. Boedecker da Fundação Getulio Vargas - SP

BARBARA WOEGERBAUER

COMPENSATION EFFECTS IN A SOCIALLY RESPONSIBLE INVESTMENT CONTEXT

Thesis presented to Escola de Administração de Empresas de São Paulo of Fundação Getulio Vargas, as a requirement to obtain the title of Master in International Management (MPGI).

Knowledge Field: Economia e Finanças Internacionais

Approval Date
22/11/2019

Committee members:

Prof. Dr. Rafael Schiozer

Prof. Dr.Jéfferson Colombo

Prof. Dr. Marcio Gabrielli

ABSTRACT

A central idea of Socially Responsible Investing (SRI) is that investors accept a loss in financial returns if they are compensated by non-financial utility that they derive from their sustainable investment. This study investigates the idea of financial penalty in SRI by examining return and risk patterns of more vs. less sustainable assets in Europe, Japan and the United States. The four distinct ESG scores provided by Bloomberg (ESG composite score, environmental, social, governance sub-scores) are used in this thesis to proxy sustainability. Responsible and irresponsible portfolios are built by grouping assets based on their scores. Portfolios are constructed separately for the 4 distinct ESG dimensions and for 3 geographies. The resulting 12 investment segments are analyzed to make a statement between a potential financial penalty in the sense of a suboptimal return and risk profile. This study finds that, in most investment segments, there is nothing to be lost from investing in the responsible portfolios compared to investing in the irresponsible portfolios, but that ESG-rated assets tend to outperform the market in general. It follows the notion of compensation effects in this specific approach is widely misleading and it appears that investing good simultaneously means investing really well.

KEY WORDS: Socially Responsible Investing, SRI, ESG Investing, Sustainability

RESUMO

Uma idéia central do Investimento Socialmente Responsável é que os investidores aceitem uma perda nos retornos financeiros se forem compensados pela utilidade não financeira que derivam de seu investimento sustentável. Este estudo investiga a idéia de penalidade financeira, examinando os padrões de retorno e risco de ativos mais versus menos sustentáveis na Europa, Japão e Estados Unidos. As quatro pontuações ESG distintas fornecidas pela Bloomberg (pontuação composta ESG, sub-pontuações ambientais, sociais e de governança) são usadas nesta tese para proxy da sustentabilidade. Portfólios responsáveis e irresponsáveis são criados agrupando ativos com base em suas pontuações. Os portfólios são construídos separadamente para as 4 dimensões ESG distintas e para 3 geografias. Os 12 segmentos de investimento resultantes são analisados para fazer uma declaração entre uma penalidade financeira potencial no sentido de um retorno subótimo e um perfil de risco. Este estudo constata que, na maioria dos segmentos de investimento, não há nada a perder ao investir nas carteiras responsáveis em comparação com as carteiras irresponsáveis, mas que os ativos com classificação ESG tendem a superar o mercado em geral. Segue-se que a noção de efeitos de compensação nessa abordagem específica é amplamente enganosa e parece que investir bem simultaneamente significa investir muito bem.

PALAVRAS-CHAVE: Investimento Socialmente Responsável, ESG, Sustentabilidade

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

CAPM	Capital Asset Pricing Model
cf.	confer
CRSP	Center for Research in Security Prices
e.g.	*(lat.) exempli gratia*, for example
ENV	Bloomberg Environmental Disclosure Score
ESG	Environmental, Social and Governance
Eurosif	European Sustainable Investment Forum
GOV	Bloomberg Governance Disclosure Score
GSIA	Global Sustainable Investment Alliance
i.e.	*(lat.) id est*, that is
JSIF	Japan Sustainable Investment Forum
KLD	Kinder, Lydenberg, and Domini Research & Analytics
PRI	Principles for Responsible Investment
SOC	Bloomberg Social Disclosure Score
SRI	Socially Responsible Investments
US SIF	US Forum for Sustainable and Responsible Investment
USD	US Dollar

1. Introduction

1.1 Purpose

Considering the recent growth in Socially Responsible Investments (SRI), mixing money and morality appears to be a major trend in the investment world. According to estimations by the Global Sustainable Investment Alliance (GSIA, 2019, p. 8), the growth of SRI assets amounted to 34% in the last two years, resulting in an overall SRI volume of 30.7 trillion USD in 2018 in 5 major markets (Canada, Europe, Japan, Australia/New Zealand, United States). Whereas the volume of SRI assets as percentage of total managed assets varies between 20% (Japan) and 65% (Australia/New Zealand) in the major SRI markets, the impressive growth numbers provide no reason to believe that the trend will reverse anytime soon. The burning question is what can be held accountable for this movement; From an idealistic perspective, it is likely that various recent events, most notably the 2008 financial crisis and the subsequent cry for the "age of responsibility", or the increasing awareness for humanity's responsibility for climate change, have changed the mindset of solely profit-oriented investors and have created a deeper sense of responsibility in the financial world (Connaker & Madsbjerg, 2019; GSIA, 2019, p. 14; Puaschunder, 2017, p. 210). According to this logic, it might have become more important for investors "to do good", to the point that they accept a less-than-optimal risk-return profile for the sake of investing in sustainable assets, or at least refrain from investing in non-sustainable assets or "sin stocks". Non-financial utility could have become a third criterion besides the two paramount investment criteria in finance: risk and return (Derwall, Koedijk, & ter Horst, 2011, p. 2137). While it could indeed be the case that these compensation effects exist and non-financial utility, derived in various ways from investing sustainably, makes up for a comparatively bad risk-return profile, it could also be possible that SRI assets actually do provide an optimal risk-return profile or at least one that is comparable to conventional investments. With any non-financial utility being "the cherry on the top" of good financial returns (i.e., no compensation effect), the growth in SRI assets might also be explained by the fact that investors have realized that focusing on doing good means simultaneously investing really well.

1.2 Research Questions, Methodology and Results

Using different methodologies, the financial pay-off of sustainable investments has been evaluated on fund-level (cf. Bauer, Koedijk, & Otten, 2005), index-level (cf. Belghitar, Clark, & Deshmukh, 2014) and on firm-level (cf. Auer & Schuhmacher, 2016; Halbritter & Dorfleitner, 2015; Kempf & Osthoff, 2007; Oikonomou, I., Brammer, Brooks, & Pavelin, 2012). This paper contributes to the subject of SRI performance by revisiting the evidence from the perspective of investors in the three main SRI markets US, Europe and Japan

(GSIA, 2019, p. 8) and utilizing Bloomberg's ESG ratings. From a theoretical point of view, does non-financial utility exist in an SRI context, and how can it potentially be reflected in stock returns? Moving away from the traditional investor who is solely interested in the risk and return of an investment, an SRI investor is introduced who chooses stocks based on ESG (environmental, social and governance) ratings. Bearing in mind the increasing importance of SRI assets as seen by their growth in volume, and assuming that SRI investors gain non-financial utility from higher ESG scores, is there a pattern that suggest that more sustainable assets are less attractive financially, i.e. that higher ESG scores compensate for lower financial returns and/or higher risk? Even if such a compensation effect might not be visible in the entire ESG market, can it be observed in different degrees for specific ESG scores or certain geographies?

The relation between ESG scores, returns and risk is tested by forming responsible and irresponsible portfolios containing assets with the highest and lowest ESG scores, respectively. A difference portfolio is created to check if a potential relationship between ESG ratings and the financial performance can be exploited by a long-short strategy of a difference portfolio (long in irresponsible, short in responsible assets), controlling for risk factors that typically explain stock returns. Risk is assessed both on portfolio level, applying an Adjusted Sharpe Ratio that takes into account higher moments of risk, and on stock level via correlation analysis. The process is done for all 12 investment segments, resulting from evaluating 3 geographies and 4 distinct ESG scores (ESG composite score and Environmental, Social and Governance subscore).

The results show that portfolios consisting of assets that are rated with ESG scores, no matter if these scores are low or high, tend to outperform the market. The performance difference between responsible and irresponsible portfolios is secondary, as the results of the difference portfolios show a lack of significance in the estimated alphas. Only in the US *ESG* and *Governance* dimension, SRI investors lose utility in terms of risk and return when investing in the responsible portfolios. It follows that the notion of compensation effects is mostly misleading in the sense that the non-financial utility to be gained from the sustainable nature of ESG-rated assets comes on top of the financial performance.

1.3 Outline

The thesis is structured as follows. Chapter 2 provides an overview of the development and the current status of the SRI industry. Moreover, the idea of non-financial utility in an SRI context and empirical results of related studies are discussed to derive the hypotheses for this thesis. In chapter 3, the methodology and data are explained in more detail, and descriptive statistics of the data set are provided. Chapter 4 reports the results, which are discussed in greater detail in chapter 5. Chapter 6 concludes.

2. Research Context

2.1 SRI - A Melting Pot of Various Sustainable Investment Strategies

According to the European Sustainable Investment Forum (Eurosif), SRI can be defined as a "long-term oriented investment approach which integrates ESG factors in the research, analysis and selection process of securities within an investment portfolio. It combines fundamental analysis and engagement with an evaluation of ESG factors in order to better capture long-term returns for investors, and to benefit society by influencing the behavior of companies" (GSIA, 2019, p. 21). To reach these two major goals of socially responsible investing – financial returns and societal benefits – the main strategies, as measured by global assets under management, continue to be *negative screening* and *ESG integration*.

The first strategy entails the exclusion of certain controversial sectors (e.g. alcohol, tobacco, gambling, military) or even single firms from portfolios; the respective assets are commonly referred to as "sin stocks" as they "make money from human vice" (Blitz & Fabozzi, 2017, p. 1; Kempf & Osthoff, 2007, p. 909). *ESG integration* means the active integration of ESG criteria into financial analyses by portfolio managers (GSIA, 2019, p. 21). More specifically, the *positive screening* describes a strategy in which assets with the highest ESG scores are combined in a portfolio; if these assets stem from different sectors, the approach is *called best-in class*. Other than the negative screening approach, this strategy does not exclude assets from critical sectors as long as they yield a high ESG score in this given sector (Kempf & Osthoff, 2007, p. 909). However, various other approaches to combine financial and non-financial goals in investment management can be subsumed under the term SRI, including rather vague concepts such as *shareholder action*, defined as "the use of shareholder power to influence corporate behavior", and *sustainability themed investing*, defined as the "investment in themes or assets specifically related to sustainability" (GSIA, 2019, p. 7). In a study examining these different strategies to implement ESG, Gibson, Krueger and Matos (2019, pp. 17–18) find that they differ significantly across geographies. As an example, European SRI investors tend to implement screening procedures much more than US investors. SRI is a broad concept not only with regards to strategies; also when directly asking SRI investors about their understanding of SRI and about their ethical, societal or religious goals, the definition of social responsibility varies greatly between people, amongst others depending on their social, economic and religious background, as shown by Statman (2008, pp. 4–6) in his quiet conversation with different SRI investors.

The range of understandings, investment approaches and differences in their concreteness makes SRI seem like a fuzzy concept when conducting literature review. Indeed, as shown in section 2.3 which gives an overview of the on financial performance of SRI according to

different studies, the results appear even contradicting – a reminder that SRI does not refer to one single investment strategy, and a reminder to establish a clear definition when doing own research.

Being aware of various definitions and strategies, the GSIA allows for different terminology describing the same or closely related investment approaches (sustainable investing, impact investing, ESG investing etc.) in its latest report, published in May 2019 and covering updates from the world's largest GSIA members including the US Forum for Sustainable and Responsible Investment (US SIF), Eurosif and the Japan Sustainable Investment Forum (JSIF). The alliance leaves no doubt that the SRI market has overall been growing in size and importance in all parts of the world, even if with different speed. In the US, Europe and Japan alone, SRI assets amounted to USD 28.25 trillion in 2018, with compound annual growth rates amounting to 6% (Europe), 16% (US) and impressive 308% (Japan) between 2014-2018 (GSIA, 2019, p. 8). The growth rates need to be considered in relation with the total volume of professionally managed assets to better estimate the relevance of SRI in different regions: The percentage of SRI is at 48.8% in Europe, at 27.5% in the US and at 18.3% in Japan. Also, in view of the different growth rates, it is worth looking at the global SRI market, where Europe has been holding around 50%, whereas Japan at 7% has a lot more catch up potential. Moreover, different standards regarding SRI investing and consequent terminological restrictions, notably in Europe to avoid "greenwashing", have led to divergences in defining SRI assets and thus, to great disparities in official growth rate. Not only the growth rates of the GSIA suggest that the success story of SRI is anywhere close to coming to an end. The Principles for Responsible Investment (PRI) Association, a UN-supported network which aims at bringing SRI investors together, highlights high growth numbers of signatories of the initiative, a strong increase in asset owner reporting and further efforts to strengthen the relationship among SRI investors to reach the goals of the ambitious 10-year blueprint towards more sustainable markets and a more prosperous world for everyone in line with the UN Sustainable Development Goals (PRI, 2018).

2.2 Non-Financial Utility as a Compensation Effect in SRI

2.2.1 Utility in SRI vs. in the Classic Finance Theory

Classic finance theory suggests that investment decisions are solely based on profit maximization and an optimization of the risk and return profile based on Markowitz' (1952) mean-variance framework. In this view, investors are rational and balance risk and return in such a way that higher risk needs to be compensated by higher return. Put more technically, investors choose their optimal investments based on return distribution and covariation, as well as their risk appetite, and the optimal risk-return profile can be

achieved by means of diversification across assets. The well-known Capital Asset Pricing Model (CAPM) by Sharpe (1964) is based on these central features.

There are several strands of research that have challenged these assumptions of classic finance. Without referring exclusively to SRI, Fama and French (2007, pp. 668–669) emphasize the low explanatory power of the CAPM model with regards to stock returns, and explain that this type of asset pricing models fails to acknowledge that investment goods are also considered consumption goods. The researchers stress the fact that investors are not only interested in the consumption of the expected financial payoff of their assets, but of the nature of the asset itself, as indicated by the fact that investors hold certain assets (e.g. of their home market or of their employer) to an extent that cannot be justified by the expected payoff of these assets alone.

The notion of investors deriving utility from aspects additionally to risk and return is central to the literature about SRI: Statman (2005, p. 31) dismisses the idea of purely rational investors, who are guided by utilitarian benefits (i.e. risk and return) and "do not care about such characteristics as (…) the social responsibility of their companies unless those characteristics affect the risk and expected return of their overall portfolio". Instead, he introduces "normal" investors who are prone to cognitive biases and emotions, based on "expressing benefits", expressing their values, tastes, wealth, and social class, to others and themselves. Derwall, Koedijk and ter Horst (2011, p. 1) call this type of investor "value-driven", with the central feature that "social and personal values instead of financial considerations are the basis for the decision to integrate corporate social responsibility criteria into investment decisions". Puaschunder (2017, p. 228 ff.) who studies socio-psychological motives of SRI investors, even allows SRI investors to derive non-financial utility from the transparency or information closure of SRI assets itself.

No matter the denotation, the inclusion of non-financial utility in the allocation considerations of investors has central consequences: First, it questions the claim of universal validity of the mean-variance theory, according to which investors with a certain risk appetite act homogenously in their optimization (Beal, Goyen, & Philips, 2005, p. 3). Even more importantly, a potential underperformance in terms of risk and return is put into perspective, as long as the value-driven investor is compensated by the third dimension (non-financial utility) the asset provides (Derwall et al., 2011, p. 1).

2.2.2 Different Forms of Non-Financial Utility in SRI

There has been plenty of research about the exact nature of the non-financial utility derived from investing responsibly, which might act as a compensation effect for a suboptimal risk-return profile or SRI assets. Most of them can be broadly grouped in two

categories: Intrinsic motivation for genuine positive social change, and social motives (creating a desirable reputation by signaling social responsibility towards the investor herself and towards the peer group). Beal, Goyen and Philips (2005) stick exactly to this categorization in their theoretical study on utility models where they identify the two main motivations for SRI, besides financial returns, to be non-wealth returns and social change. Non-wealth return refers to the utility derived from the happy feeling of investing in a good cause that is also considered desirable by the peer group, no matter if the investment is genuinely ethical or not. However, these non-wealth returns are not a strong enough investment motive by itself, but they need to be accompanied by substantial financial return expectations.

The second non-financial motivation for SRI investors is social change, meaning the genuine expectation that their investment in social issues will actually improve the state of the world. However, since genuine social change is both difficult to measure and hard to achieve especially for single investors, this motive for investing in SRI assets is thought to be inferior and only plays a role in combination with financial returns and non-wealth returns. The distinction between an investor's wish to actually achieve positive social change, and the mere effort to "gain peace of mind and consistency between their values and their investments" ("expressive benefits" or again, "non-wealth returns" in the spirit of Beal et al.) by investing in SRI assets is also emphasized by Statman (2008, p. 9).

Regarding this dichotomy of rather altruistic and egoistic motives, it is important to consider the design of studies which empirically test the nature of non-financial utility: Derwall et al. (2011, p. 2140) highlight the fact that investors tend to give socially desirable answers, thus it is crucial to use methodologies which do not only include survey data or are able to conceal the nature of the study. A study by Bauer and Smeets (2015) who use survey data of investors that make actual investment decisions emphasizes especially the social motives of SRI investors: As a widely used concept in economics that drives consumer behavior, social identification is found to be also relevant in an SRI context when investors aim at creating their pro-social image by investing in an ethical manner. Positive social identification strongly correlates not only with the fraction invested in SRI, but also with the overall portfolio size (Bauer & Smeets, 2015, p. 128). The categorization of intrinsic motivation and social motives is further supported in a study by Riedl and Smeets (2017): Using administrative investment data, survey data and incentivized experiments, the researchers identify intrinsic social preferences and social signaling to play a significant role in investment decisions besides the financial aspects of risk and return: Less selfish investors with social motives which bring no material benefit to themselves, as well as investors who talk a lot about their investments are significantly more likely to choose SRI instead of conventional funds. However, the amount invested in SRI funds is rather related

to the overall portfolio size of the investors for diversification reasons, whereas intrinsic social motivation and signaling social motives are only relevant for the initial investment decision (Riedl & Smeets, 2017, p. 2522 ff.). Nonetheless they conclude that the effect of SRI investors on asset prices will grow in the future as the fraction of SRI in the market continues to increase (p. 2533). Puaschunder (2017, p. 224) emphasizes the eternal human pursuit of innovation which can be linked to the intrinsic motivation for positive social change: "SRI allows investors to reward societal progress and innovatively tackle social and environmental concerns". Last but not least, very expressive results in favor of non-financial utility and especially of intrinsic motivation for social change are found in a survey-based study by Brodback, Guenster and Mezger (2018), who establish a link between altruism and the importance of SRI. Acknowledging that financial and non-financial motives coexist in the investment world, and dividing potential investors into egoistic and altruistic ones, they find that the altruistic investors are more inclined towards investing in SRI when the actual impact is thought to be big, whereas egoistic investors are triggered by higher returns. The reason for altruistic investors to consider the actual impact in their investment decision is a strong moral sense of duty to contribute when the "perceived SRI effectiveness" is high (p. 16). Linking these findings to the interplay between financial and non-financial utility, the study finds an attractive financial return even decreases altruistic investors' propensity to invest in SRI (p. 18). The intuitive logic behind "perceived SRI effectiveness" as a reinforcing factor behind social motives to invest in SRI is also subject of other papers. Notably, in a research project including experimental studies about socially responsible behavior not limited to the investment sphere, Cojuharenco, Cornelissen and Karelaia (2016, pp. 79–81) find a strong relationship between the perceived effectiveness and socially responsible behavior, but also a link between a sense of "connectedness" between and perceived effectiveness. Socially inclined people feeling more connected to likeminded people expect their consumption choices to be more impactful, which again leads to more socially responsible behavior. Whereas SRI is considered a method to effectively reach positive social change using the logics of financial markets, it should not be unmentioned that SRI investments are not considered a substitute to charity. Riedl and Smeets (2017, pp. 2532–2533) reject the hypothesis that investors might consider SRI as a more effective and efficient way than charitable donations by testing the inclination of SRI investors to donate, as opposed to non-SRI investors.

Finally, regulatory aspects should not remain unmentioned in the context of financial and non-financial motives in SRI. Gibson et al. (2019, p. 1) emphasize that SRI has moved from being a trend of purely norm-constricted investors to reaching "mainstream institutional investors". There is also empirical evidence suggesting that institutional investors, such as pension funds, universities or insurance companies, are heavily invested in SRI funds (75% of sustainable assets are held by institutional investors, according to the GSIA (2019, p.

12)), or that they shun "sin stocks" (Hong & Kacperczyk, 2009, pp. 23–26). On one hand, this type of investors faces considerate public exposure and scrutiny, which strengthens the argument of social motives (building a desirable reputation through signaling) as a source for non-financial utility. On the other hand, these "norm-constrained investors" might also be legally restricted in their investment strategies, which – if infringement goes beyond the loss of desirable reputation and is related to financial penalties – should be logically attributed to the financial than the non-financial utility aspects. According to the PRI, there has been a surge in policy interventions since the 1990ies, but mostly since the 2007/08 financial crisis. These policies typically evolve around regulations for central asset owners, notably pension funds and key state-owned asset owners, an improved interaction between investors and investee companies for better governance and shareholder protection, and the encouragement of corporate disclosure regarding ESG issues (PRI, 2019). For example, two major institutional asset owners in Japan (the Government Pension Investment Fund in 2015 and the Pension Fund Association in 2016) recently became signatories to the PRI, with the aim of promoting policies based on these principles (GSIA, 2019, p. 15). In various European countries, pension funds and other institutional investors are required by law to disclose ESG-related aspects of their investments (Renneboog, ter Horst, & Zhang, 2007, p. 1727). Interesting in the context of regulation is the notion that certain types of investors might be restricted from SRI investing due to fiduciary restrictions, i.e. „investors' obligation (...) to act in the best interests of beneficiaries", which would be violated if financial goals were not sufficiently met due to a blind emphasis on the ESG-nature of investments – which is, however, not the case according to the PRI (2019). Quite the contrary, the GSIA (2019, p. 14) lists fiduciary duty even as a motive for investors to choose SRI assets, emphasizing the belief that these investments yield a satisfactory financial result.

Given the range of non-financial motives to invest in SRI, firms perceive the necessity to invest in sustainability and disclose ESG-relevant information, so as to remain an attractive investment goal for investors who put emphasis on non-financial utility in their strategy (or for investors who are norm-constricted for legal reasons). This movement suggest that the financial world has come a long way from the free-market view substantially shaped by Milton Friedman (1970), who prominently stated that business managers spending money on the social agenda of their company effectively pass on these costs, and thereby "tax", stockholders, consumers and employees, which is what only the elected government should be doing.

2.3 The Return Puzzle in SRI

Having learned that non-financial motives do exist for SRI investors, it is important to understand how they might be reflected in stock prices. The empirical results regarding the

financial performance of SRI assets are manifold and suggest, at first sight, that there is no stringent conclusion to be made about the costs or benefits of SRI investing (for an extensive literature overview of more than 2200 primary performance studies, see Friede, Busch and Bassen (2015)). However, as explained in the end of this section, many differences in study results are due to definitional and methodological distinctions, making it difficult to make a general statement about the over- or underperformance of SRI investments. First, an overview of the three possible scenarios including underlying stock price dynamics is presented.

2.3.1 Financial Returns are Independent of Social Responsibility

The efficient market hypothesis states that „ share prices reflect all information and consistent alpha generation is impossible“ and „only inside information can result in outsized risk-adjusted returns“ (Investopedia, 2019). If the investment theory holds with respect to SRI, there should be no return differences between responsible (high ESG) and irresponsible (low ESG) companies, i.e. the alpha of a long-short strategy should not be significantly different from zero. As a prerequisite for the costs and benefits of ESG to be fully and truly reflected in stock prices, sufficient information about the pricing relevance of ESG must be available to all investors (Manescu, 2011, p. 97).

Using an international data set of mutual funds, Bauer, Koedijk and Otten (2005) find no evidence of significant differences in risk-adjusted returns between ethical and conventional funds. A study by Belghitar, Clark & Deshmukh (2014, p. 59) using SRI indices confirms the theory that there is no return difference between responsible and irresponsible investments. Blitz and Fabozzi (2017) conduct a study about the return characteristic of sin stocks. These assets are therefore often shunned in a negative screening process by investors, which would explain a superior return due to systematic underpricing and the existence of specific investment vehicles for investors who seek to monetize their lack of morals in an investment context. Controlling for well-known risk factors size, value, momentum, as well as investment and profitability, their findings show that the inclusion of the two new factors not only increases the explanatory power of the model – as measured by R^2 – but also decreases the alphas below significance level. They conclude that no premium exists which could be attributed to the sinful character of these stocks (e.g. by means of a premium for reputational issues); moreover, the sin stocks' superior financial can be fully explained by their exposure to established risk factors (Blitz & Fabozzi, 2017, pp. 4–5).

2.3.2 Doing Good while Investing Well

Statements by institutions like the PRI (2019) or the (GSIA, 2019, p. 20) regarding the superior performance of "good" stocks obviously needs to be taken with caution, given the agenda of these organizations to strengthen the standing and deepening the impact of SRI in the investment world. However, in their meta-analysis of previous studies, also Friede et al. (2015, p. 217ff.) find a positive relationship between measures of sustainability and financial performance in most of the research conducted on the financial performance of SRI – across asset classes and geographies. As an example for a performance study using portfolios based on ESG criteria, Kempf and Osthoff (2007, pp. 914–918) implement a strategy that is long in high-ESG assets and short in low-ESG assets (both through *positive screening* and *best-in-class screening*) and find, controlling for various risk factors, that this approach yields a significant positive abnormal alpha.

Several explanation attempts for the superior performance exist, mostly relating to the pricing relevance of the investors' perception of costs and benefits of a firm's social responsibility (Manescu, 2011, pp. 97–98). The costs of SRI generally concern a firm's investments in sustainability to yield higher ESG scores and to publish and advertise these efforts accordingly, at the expense of shareholder value, whereas the benefits notably materialize in the minimization of litigation risks, reputational risk etc. (Hong & Kacperczyk, 2009, p. 17). – that is not to make a universal statement about bottom line of SRI's costs and benefits. However, if benefits outweigh the costs but investors overestimate the costs or underestimate the benefits, it will result in a mispricing of ESG in the sense that positive abnormal returns ("positive earning surprises") in high ESG investments occur (Huppé, 2011, pp. 45–46). According to Derwall et al. (2011, p. 2139) there are several good reasons why investors would underestimate the benefits of SRI ("errors in expectation"), leading to positive earning surprises: The lack of tools for investors to completely grasp the multidimensional concept of social responsibility and its impact on firm value, the failure of accounting standards to capture these benefits which are often intangible, and the tendency of financial markets to pay more attention to negative than to positive news. However, since markets eventually learn about the positive values of SRI, abnormal returns should disappear over time. De and Clayman establish a link between ESG Scores, volatility and returns. They find that ESG scores are related to lower volatility, and that this lower volatility is associated with higher returns, however, ESG scores can also be considered an independent contributor to returns (2015, pp. 50–51).

2.3.3 Morality Comes at a Price

There are several theories supporting the idea that investing in sustainability comes at the cost of financial performance as measured by stock returns. The technical argumentation

against the superiority of SRI is based on Markovitz' mean-variance framework and states that SRI investments can only underperform conventional assets in the long run because they are, by definition, a subset of the market portfolio and therefore restrict diversification opportunities (Derwall, 2007, p. 12). Another theory takes up the aforementioned economic argument interplay of costs and benefits related to SRI: If costs turn out to be higher than benefits but investors underestimate these costs and simultaneously overestimate the benefits, they create upwards pressure on SRI asset prices, yielding in lower returns. A third logic refers back to the reduction of certain risks by responsible firms. Low ESG companies on the other hand still face these risks and thus, investors buying low ESG assets or even sin stocks" yield a higher return because they require to be compensated for this additional "sin" risk factor (Manescu, 2011, p. 97).

The argument of social responsibility negatively impacting stock prices can also be based on the "discriminatory taste" concept in economics, as articulated in an early model by Becker (1957): If community norms shape discriminatory tastes of agents and result, for example, in the decision not to interact with certain people, these discriminatory tastes might result in higher financial costs for the agents. In the context of SRI, Heinkel, Kraus and Zechner (2001) take Becker's discriminatory taste argument one step further and develop a theoretical model showing that a sufficient number of investors restricted by community norms (in the case of their model, environmental norms) can influence stock prices in the market: Stocks that are shunned by norm-constrained investors in the market will be cheaper and bring higher financial returns than acceptable stocks; the dynamics are simply based on the law of supply and demand, or, more technically, on the lack of risk-sharing between the (few) investors that still invest in critical stocks in a norm-constrained environment. As opposed to the economic argument, the interplay of costs and benefits and the bottom line only play a secondary role in this logic. As Derwall et al. (2011, p. 2139) explain, two assumptions must hold in order for the discriminatory taste concept to have a negative influence on SRI stock returns: SRI investors must be "value-driven", i.e. they must derive non-financial utility from investing in SRI, and there must be a significant number of them. Similarly to Heinkel et al., Fama and French (2007, pp. 675–677) derive an asset pricing model that considers tastes for assets as consumption goods (e.g. SRI), which works similarly as an equilibrium model that includes traders with "misinformed beliefs". They explain that the existence of asset tastes can significantly distort expected asset returns, if the respective investors account for a large enough portion of investors and their positions deviate much from the market portfolio.

The claim is confirmed by a variety of studies including Hong and Kacperczyk (2009, pp. 27–31), who find that norm-constrained institutional investors, such as pension funds, universities or insurance companies, face a higher cost of capital by shunning sin stocks.

Derwall et al. (2011, p. 2141) confirm the scenario especially with respect to negative screening strategies. On the other hand, a study by Borgers, Derwall, Koedijk and ter Horst (2015, p. 28) detects positive abnormal returns for mutual finds that have a higher percentage of sin stocks.

2.3.3 Attempts to Solve the Puzzle

Before dismissing previous studies on the financial returns of SRI as non-significant due to the ambiguous results, a few aspects should be considered. The following paragraph also serves to justify some of the methodological aspects chosen in this paper to establish a link between the financial performance of SRI assets and non-financial utility as a compensation effect.

First, as mentioned in 2.1, SRI is a very broad concept, and so are the possibilities to proxy the social responsibility of an investment. Researchers typically use ESG scores from various providers, mostly Thomson Reuters Asset4 or Kinder, Lydenberg, and Domini Research & Analytics (KLD), less often Sustainalytics or Bloomberg, each comprising different geographic areas, numbers of companies covered, underlying criteria to calculate the scores, subcomponents and time spans. To account for these differences and to create consistency, Halbrittner and Dorfleitner (2015) use three of the most commonly used proxies for sustainability in their performance study - ESG data provided by ASSET4, Bloomberg and KLD. They confirm the suspicion that some of their results strongly depend in their direction and significance on the rating provider (p. 34).

Also the strategy to include social responsibility aspects in a portfolio is decisive: As shown in section 2.1, the approaches vary widely and range from very concrete (e.g. negative screening) to rather abstract strategies (e.g. sustainability themed investing) and by no means imply the same methodology. For example, if investors shun “sin stocks” in order to avoid entire industries it does not imply that these stocks have a particularly low ESG score. For example, Japan Tobacco Inc. has an ESG score of close to 50 in most of the years in the sample of this study and lies well above the mean value. In other words, shunning sin stocks from a given portfolio in order to please an investor’s preference for ethical investments is a completely different, and not just “opposing” approach than positive screening i.e. choosing the best stocks based on ESG scores. Derwall et al. (2011, p. 2139) solve the return puzzle by distinguishing between “value-driven” investors, who sacrifice returns for the sake of non-financial utility obtained from SRI investors, and “profit-seeking” investors who seek to achieve superior financial performance by investing in SRI. By means of identifying investment approaches that meet the respective values and goals of both types, they find that “value-driven” investors are more likely to use negative screens, whereas “profit-seeking” investors tend to use positive screens; the first type of investors therefore lends

support to the "discriminatory taste" (or: "shunned stock") argument, the latter to the "errors in expectation" argument – which helps to explain why both SRI and sin stocks can show superior results. Since certain investment strategies involve both negative and positive screening, the respective effects of over- and underperformance might even cancel out, leading to the perception that there is no relationship between financial performance and social responsibility (p. 2145). Obviously it is difficult if not impossible to determine the fraction of "value-driven" vs. "profit-seeking" investors in a given investment[1]: As Derwall et al. (2011, p. 2143) put it with regards to their findings: "The risk-adjusted returns summarized in this section can be interpreted as evidence of expectation errors by investors, but they can also imply that the stocks of firms with strong environmental and/or social performance experience price increases following in- creased demand by (values-driven) socially responsible investors". What remains is to accept that "the effects of SRI on portfolio choices and asset prices cannot be understood without acknowledging the distinct role that each segment of the SRI movement plays in financial markets" (p. 2140).

Next, the type of study to compare SRI vs. non-SRI assets is relevant: Some researchers use mutual funds (e.g. Bauer et al., 2005; Renneboog et al., 2007) in their studies, others use portfolio approaches (e.g. Auer & Schuhmacher, 2016; Halbritter & Dorfleitner, 2015), others use indices (e.g. Belghitar et al., 2014). Especially mutual fund studies have received much criticism and are deemed inappropriate for a performance analysis of ESG assets. Two reasons account for this shortcoming: First, SRI mutual funds are not really worthy of their name, as they do indeed exclude the least ethical assets, however, on average do not hold more sustainable assets than conventional funds (Utz & Wimmer, 2014, pp. 79–81). Secondly, irrespective of the level of sustainability reflected in an SRI mutual fund, there are many important factors other than the ethicalness of the fund that play a role for the performance and are very hard to control for. Indeed, the performance of an SRI fund depends both on the components' firm-level performance as well as the fund's management performance (Kempf & Osthoff, 2007, p. 909). Similar to the caveats of mutual fund studies are those related to the institutional SRI investors: Gibson et al. (2019, p. 18), who calculate value-weighted average portfolio-level ESG scores of institutional investors who have joined the PRI, find that institutional investors associated with the PRI network do not necessarily exhibit better ESG outcomes. While there are some time- and geography-specific differences, it remains to acknowledge that any study that automatically assumes a higher level of responsibility from the denotation of a socially responsible investor, institutional investor or mutual fund, needs to be taken with a grain of salt.

[1] In mutual fund studies, this task is sometimes achieved by looking at the volatility of fund flows especially after periods of negative performance, e.g. (Bollen, 2007).

Moreover, time horizon seems to play a role as SRI performance is expected to change over time, be it due to awareness and demand or learning effects: According to the study of mutual funds by Bauer et al. (Bauer et al., 2005, p. 1765), these funds "underwent a catching-up phase, possibly due to learning, before delivering financial returns similar to those of conventional mutual funds". Whereas some researchers account for the possibility of performance differences over time by splitting there sample into sub-periods (Hong & Kacperczyk, 2009; Huppé, 2011; Kempf & Osthoff, 2007), others do not.

Finally, attention should be drawn to Belghitar et al. (2014, pp. 54–55) who state that higher moments of risk play an important role in the utility maximization of investors, however, they are neglected in most performance studies which are based on mean-variance frameworks such as the CAPM. Taking up the criticism of these frameworks as explained in 2.2, they do not attempt to include non-financial utility in asset pricing models but focus on higher moments of risk in their performance evaluation so as to create a move comprehensive assessment of financial-utility. Testing responsible and irresponsible indices with Marginal Conditional Stochastic Dominance, they find that there is a price to be paid for socially responsible investments in form of lower skewness and higher kurtosis, both of which are unfavorable for a risk-adverse investor with a concave utility function (p. 60). This finding somewhat puts into perspective their first findings of non-significant performance differences between SRI and non-SRI indices based on a mean-variance framework.

2.4 Legitimation and Hypotheses

As outlined in the previous section, various studies have made clear that non-financial utility exists in an SRI context based on various self-centered or altruistic motives, and that it might compensate "value-driven" SRI investors for lower financial results. In other words, if non-financial utility derived from the nature of sustainable investment is acknowledged in the SRI investors' optimization process, they might accept a suboptimal return/risk profile for the sake of investing responsibly. The question that remains is: Do SRI investors actually have to accept lower financial results, "value-driven" or not? If not, the notion of compensation effects appears misleading in the sense that SRI investors get both ESG scores and good returns – which would satisfy both types of investors in the spirit of Derwall et al. (2011). Therefore, the primary goal is to evaluate if there are return and risk patterns that suggest financial penalty, i.e. investors who pick stocks based on their level of sustainability suffer from lower financial returns and/or higher but they are compensated by non-financial utility derived from the promise of the sustainability scores.

The most important take-away from the previous section is that SRI studies can only make a statement from a specific angle, due to the wide range of study designs and methodological possibilities. The analysis of financial penalty in this paper employs Bloomberg's ESG scores as a proxy of a company's sustainability. Despite being a comparatively young provider of ESG data, the number of available scores has risen a lot in previous years for companies worldwide, thus allowing for the separate consideration of three distinct markets. Moreover, Bloomberg provides separate subscores for the three dimensions of sustainability, i.e. environment, social and governance, besides a composite ESG score. The structure and international coverage of Bloomberg's ESG scores thus supports the central contribution of this paper to the literature on SRI: Bloomberg's ESG scores allow for the segmentation of a comprehensive data set into 12 investment segments to make statements about compensation effects in three different geographies and for four distinctive ESG measurements, as outlined below.

Based on Manescu's (2011) discriminatory taste argument and the insights about an only temporary "errors in expectation" hypothesis by Derwall et al. (2011), but mostly based on theoretical explanations of non-financial utility which has already been proven to exist in survey- and experiments-based studies, there is good reason to believe that high ESG assets yield a lower return than low ESG assets due to the simple laws of supply and demand. To test this logic, the first hypothesis which this paper hopes to reject is:

H1: There is no difference in the returns, adjusted for common risk factors, between socially responsible and irresponsible stocks as measured by their ESG scores.

Financial penalty does not only concern risk-adjusted returns which are analyzed in common asset pricing models, as emphasized by Belghitar et al. (2014). Especially since the topic of higher moments of risk is not very present in the numerous performance study, it is taken up in this paper to find out if the present methodology supports the idea of Belghitar et al. (2014). Expecting to see that financial utility is lost in form of negative skewness or fat tails in SRI asset, the second hypothesis that can hopefully be rejected is:

H2: There is no significant correlation between socially responsible stocks and higher risk as expressed by negative skewness and excess kurtosis.

The last two hypotheses evolve around the segmentation into 12 investment segments. Most studies about social responsibility and financial performance evolve around the US; one reason, according to De and Clayman (2015, p. 52) is the broad range of ESG standards and differences in market perception of the importance of ESG factors across regions. But it is exactly this assumption that makes it interesting to compare possible compensation

effects in form of risk and return patterns across geographies. This could be the case if two conditions were met: First, there is indeed a difference in importance of sustainability-related issues across countries. For example, as stated earlier, most ESG assets are held in Europe according to the GSIA, and a possible explanation could be increased awareness about sustainability issues, stronger restrictions for institutional investors or higher litigation risk for non-compliance. If the discriminatory price argument is true, then law and demand would say that compensation effects are most visible here. The second condition says that SRI investors only invest in domestic equities. With reference to a study by Salaber (2007) who used country-specific segmentation in Europe to make statements about SRI performance, Derwall et al. (2011, p. 2141) emphasize that the latter is a very strong assumption, on top of the assumption that the understanding of ESG is the same across geographies. The latter problem is alleviated in this study through the consistent use of a sustainability measure from only one provider – Bloomberg – which does not allow for different interpretations in different geographies. The assumption of exclusively domestic investing, however, is also central in this study. From a methodological perspective, it is justified by its acceptance in other studies; from an empirical perspective, broad evidences for a so-called "home bias" in investing make it at least plausible to some extent.

To look into geographical differences in more detail, the hypothesis says:

H3: There are no differences in in risk and return patterns between Europe, Japan and the United States.

While many studies suggest that SRI investors have a homogenous preference for sustainable investments, this study allows them to draw different levels of non-financial utility for each of the sustainability dimensions, as reflected by the fourth hypothesis:

H4: There are no differences in risk and return patterns for investments in distinctive ESG dimensions.

Three different model specifications are used to check potential alphas for persistence: Besides the CAPM and the Carhart 4-factor model, a 6-factor model is employed to account for the insights of Blitz and Fabozzi (2017) and see if their findings on the explanation of sin stocks can be transferred to low- vs. high-ESG stocks.

The details on the methodology to test the four hypotheses are described in the following section.

3. Research Methodology and Data

3.1 Data Description

The period 31/12/2006 to 31/05/2017 is considered.

ESG scores

In this study, the social responsibility of firms is proxied by Bloomberg's ESG scores. According to their own service description, "Bloomberg provides ESG data on more than 11,000 companies with over 700 environmental, social and governance indicators from company-sourced filings and third-party information, covering virtually the entire publicly investable universe" (Bloomberg Finance L.P., 2019). While comparatively young, Bloomberg's ESG database has been growing substantially since its inception in 2005 and nowadays covers more companies than established databases such as Thomson Reuters/Asset4 (established in 2002, 7000+ companies) and MSCI ESG KLD STATS (established in 1991, 3000+ companies) (Columbia University Libraries, 2009). By combining company-sourced and proprietary information and translating it into scores from 0-100 (100 being the highest), Bloomberg promises in-depth sustainability analyses and full transparency, with every data field traceable to a document (Bloomberg Finance L.P., 2014). Since Bloomberg covers companies in more than 65 countries, it is suitable for a study that covers regions outside the United States. Furthermore, by providing an overall *ESG* score (called composite ESG score[2]) as well as separate subscores for the three dimensions environment (*ENV*, e.g., waste disposal, renewable energy), social (*SOC*, e.g., diversity, community relations) and governance (*GOV*, e.g. executive compensation, independent directors), Bloomberg's disaggregate data can be used to separately consider and distinguish the three classical dimensions of sustainability.

ESG scores are updated and published by Bloomberg in the beginning of each year *y*, thus representing the score that considers ESG-relevant information the previous year *y-1.* Year *y* of data publication and portfolio construction (see 3.3.1) is used as reference year throughout the paper.

Sample description

A list of 11'726 companies worldwide that have ever disclosed ESG data is derived from Bloomberg (*BESGPRO INDEX*, *MEMB*). To limit the scope to the three geographies in

[2] *ESG* is written in italics whenever the Bloomberg ESG composite score is meant. Whenever ESG is not written in italics, the sum of all scores is meant without reference to a specific subscore.

question (US, Japan, Europe), the list is reduced to companies quoted by stock exchanges in these regions: The United States is considered a homogenous market; to make Europe as homogenous as possible and to account for the importance of individual country risk which potentially explains part of stock returns, countries which did not have prime or high medium grade rating according to the three main sovereign rating agencies over the entire observation period are excluded[3]. Also excluded are non-Euro countries (Switzerland, Norway, United Kingdom) to account for potential currency exchange volatilities. Companies listed on exchanges in the following EU countries therefore remain and are subsumed as "Europe": Austria, Belgium, Finland, France, Germany, Luxemburg and the Netherlands.

The sample likely suffers from a bias related to the fact that large corporations might have more resources to invest in sustainability and making these efforts public, including respective reports. (Drempetic, Klein, & Zwergel, 2019). At the same time, historical returns show that firms with large market capitalization typically have lower returns (Bali, Engle, & Murray, 2016, p. 152). This potential large cap bias is controlled for in the regression models that include an SMB factor as described in chapter 3.2.3.

Returns

In the next step, the companies with disclosed ESG data are checked for price data to obtain returns. In each year, companies who have disclosed an ESG data point are checked for available price data. Since 6-months lagged returns are used to avoid a look-ahead bias and the 12 ensuing monthly returns are used for portfolio return and volatility calculations (see portfolio construction in chapter 3.2.1) (cf. De & Clayman, 2015, p. 51), manual return checks start on 31/05/2007 to calculate the first return in 06/2007. ESG-rated stocks for which no sufficient return data for the 12 subsequent months can be obtained are dropped from the sample in the given year. By doing so, many companies as possible can remain in the sample.

Financial data for Europe and Japan is retrieved from Bloomberg: Monthly prices (*PX_LAST*) for return calculations as well as for calculations of measures of risk, and shares outstanding (*EQY_SH_OUT*) (in Mio), all in USD. For the United States, financial data is obtained from Bloomberg and, wherever data is not available, additionally from the Center for Research in Security Prices (CRSP), following the methodology in Bali et al. (2016, p. 105 ff.): Translating Bloomberg tickers into CRSP-compatible ones, share prices in USD

[3] http://www.worldgovernmentbonds.com/world-credit-ratings/

(*ALTPRC*[4]) and the number of outstanding shares in thousands (*SHROUT*) are obtained. *SHROUT* is divided by 1000 to also obtain outstanding shares in millions. For all geographies, market capitalization (*market cap)* for month *t* is calculated as stock price times the shares outstanding at *t* (Bali et al., 2016, p. 4):

$$market\ cap_{i,t} = price_{i,t} \times shares\ outstanding_{i,t} \quad (1)$$

The risk-free rates (rf_t) to calculate excess returns are obtained from the Kenneth R. French Data library (2019). Similarly, the common risk factors known to explain stock returns (see chapter 3.2.3) are also downloaded from the Kenneth French Data library, for each of the geographies, all in USD.

3.2 Methodology

3.2.1 ESG Portfolio Construction

Following the methodology in Bali et al. (2016, p. 39 ff.), a portfolio approach is used to investigate a relation between the dependent variable (excess returns) and sort variables (ESG scores) for each of the 12 investment segments. Forming portfolios based on ESG over- and underperformer, the univariate portfolio method aggregates the panel data sets to a single time-series dimension and enables the use of asset-pricing models and risk assessments. In comparison with linear regression, portfolio analysis does not assume a linear relationship between the two variables in question but can be used to discover any cross-sectional relation. In comparison with full sample cross-sectional regression (cf. Halbritter & Dorfleitner, 2015; Manescu, 2011), researchers mostly use a portfolio approach to investigate the relation between SRI and financial performance (cf. Belghitar et al., 2014; Derwall et al., 2011; Halbritter & Dorfleitner, 2015; Kempf & Osthoff, 2007). As it turns out when scrutinizing the ESG sample in this paper, some assets are rated with the same ESG scores over several years in a row; this observation makes the portfolio approach even more useful as assets in a certain ESG percentile are combined in a portfolio, without trying to establish a direct link within (sometimes constant over the years) ESG scores and (varying over the years) returns.

In the beginning of each year *y*, ESG scores become available on Bloomberg and refer to the disclosure of ESG-related information of the given stock in the previous year *y-1*. To avoid a look-ahead bias, which arises when data is used that has not been available at the time of the analysis, a 6-months lag is then used for the portfolio construction and

[4] *ALTPRC* instead of PRC is used to avoid data gaps if there is no trading activity on the last trading day in the given month and no bid or ask prices available (Bali et al., 2016, p. 107). The absolute value of *ALTPRC* is used to avoid negative prices which appear if there was no trading activity on the last trading day.

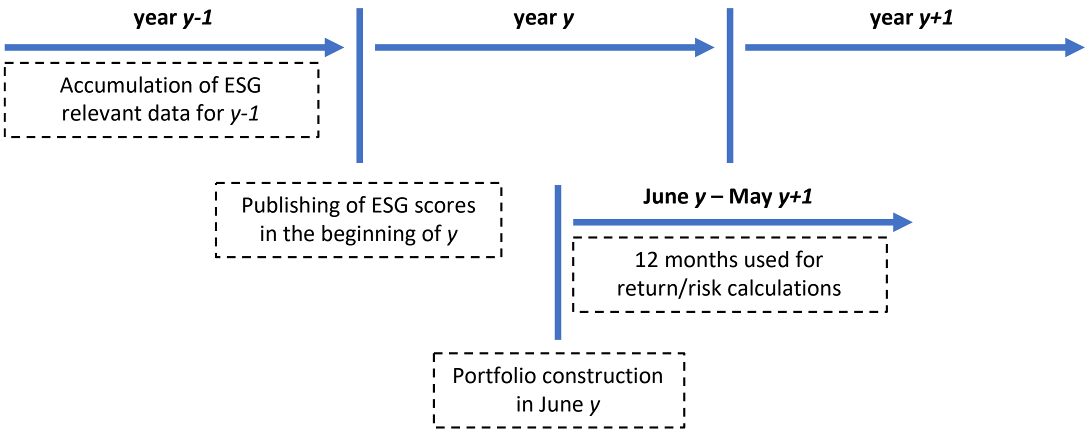
risk/return calculations (cf. De & Clayman, 2015, p. 51): In June, each stock with one or more available scores is ranked with respect to its respective ESG score(s) within its geography. For each investment segment, the sample is split into three portfolios *p* with the two breakpoints at the 30th and 70th percentile of the ESG scores: P_{70} is considered best-in-class (irrespective of industries) and called "responsible", p_{30} is considered worst-in-class and called "irresponsible". The assets in the portfolio for the mid-range is neglected in the given year. In sum, 24 portfolios are constructed: A "responsible" and an "irresponsible" portfolio, for each of the four types of ESG scores, and for each of the three geographies. Choosing the breakpoints at the 30th and 70th percentile results in a sufficient number of assets per portfolio (see table 2), especially in earlier years when fewer firms are rated, to make a statement about the true parameters and avoiding too much noise; at the same time, given the distribution of ESG scores, it is important to maintain a close to homogenous characteristic of ESG scores within the two extreme portfolios (i.e. the range of ESG scores is not too big). However, a robustness test in section 3.2.5 includes different thresholds at the 20th and 80th percentile.

When updated scores are available in the following year, the procedure is repeated: After 6 months, i.e. in June *y+1,* assets in each geography are again ranked with respect to their updated scores and assigned to the respective portfolios. By using this yearly rebalancing methodology, it is assured that each year, the 24 portfolios always contain exactly those assets with the 30% highest and 30% lowest scores, respectively. While many assets pertain in the same portfolio over years as their ESG scores do not change much, the overall number of ESG ratings increases significantly each year (see table 2), leading to a substantial turnover and diversity in the portfolios.

For the analysis of the portfolio performance and its risk measures, the returns of the stocks in the respective portfolio are calculated. 12 months of stock return data (between June *y* and May *y+1*) are used for risk and return calculations. Using this methodology, $r_{p,t}$ represent the portfolio return that would have been realized by an investor who, after obtaining ESG scores early in the year *y*, constructed the portfolios in June *y* and held the portfolios without further trading for 12 months, before the portfolios are rebalanced in June *y+1.* Figure 1 depicts the timeline of portfolio construction.

FIGURE 1: TIMELINE OF PORTFOLIO CONSTRUCTION

The 36 portfolios are constructed in the beginning of each year *y* and refer to ESG-relevant information published in year *y-1*. Year *y* is thus the reference year. The assets are held unchanged for one year, before portfolios are rebalanced based on the updated ESG scores. The return data to be used for risk and return analyses is calculated between June y and May y+1.



In line with Bali et al. (2016, p. 44), the portfolios are value-weighted as this procedure is considered to generate more realistic returns an investor would have been able to realize with the portfolio. However, this assumption is challenged in a robustness test in section 3.2.5 which uses equal-weighted portfolios. In the value-weighting process, the market capitalization of the individual stock in month t is divided by the market capitalization of the entire portfolio (the sum of all stocks' market capitalization in t), and this percentage is used to calculate the weight $w_{i,t}$ for the return of stock i in each month t:

$$w_{i,t} = \frac{\text{market cap}_{i,t}}{\sum_{i \in p} \text{market cap}_{i,t}} \tag{2}$$

From the "responsible" and "irresponsible" long-only portfolios, a difference portfolio $p_{difference}$ is formed so that the following relationship about the returns in each month t holds:

$$r_{p_{difference},t} = r_{p_{irresponsible},t} - r_{p_{responsible},t} \tag{3}$$

This difference portfolio represents a zero-cost long-short strategy that buys irresponsible stocks (low ESG scores) and sells responsible stocks (high ESG scores) to test whether this strategy can yield a positive alpha[5]. Difference portfolios are commonly used in ESG

[5] Note that an SRI investor would go long in the responsible stocks; however, this long-short strategy is merely to test the hypothesis that responsible stocks yield an inferior return i.e. irresponsible stocks yield a superior return.

portfolio analysis as a practice to detect the true cross-sectional relationship between the sorting variable (in this case, the ESG score) and the outcome variable (excess return) (Bali et al., 2016, p. 46; Halbritter & Dorfleitner, 2015, p. 26). The abnormal return of the difference portfolio (α) is focused in the analysis; if α of the difference portfolio are statistically distinguishable from zero (controlling for common risk factors in the spirit of Fama and French and Carhart) it is commonly interpreted as evidence that sorting variable (in this case, ESG scores) have some ability to predict cross-sectional variation in stock returns (Bali et al., 2016, p. 121).

In sum, we derive 36 portfolios (a high, low and difference portfolio for each investment segment, see figure 2), with a time series of returns over 11 years, i.e. 132 months (2007-2018).

FIGURE 2: OVERVIEW OF INVESTMENT SEGMENTS AND PORTFOLIOS

	ESG	ENV	SOC	GOV
EU	• Responsible • Irresponsible • Difference	• Responsible • Irresponsible • Difference	• Responsible • Irresponsible • Difference	• Responsible • Irresponsible • Difference
Japan	• Responsible • Irresponsible • Difference	• Responsible • Irresponsible • Difference	• Responsible • Irresponsible • Difference	• Responsible • Irresponsible • Difference
US	• Responsible • Irresponsible • Difference	• Responsible • Irresponsible • Difference	• Responsible • Irresponsible • Difference	• Responsible • Irresponsible • Difference

3.2.2 RETURN CALCULATIONS

Monthly returns $r_{i,t}$ are calculated for each stock i in each month t from available prices p_i:

$$r_{i,t} = \frac{p_{i,t} - p_{i,t-1}}{p_{i,t-1}} \tag{4}$$

The monthly portfolio return $r_{p,t}$ results from the sum of all value weighted ($w_{i,t}$) returns $r_{i,t}$ of the stocks that belong to the yearly rebalanced portfolio in the given month:

$$r_{p,t} = \sum_{i \in p} w_{i,t} \times r_{i,t} \tag{5}$$

Whenever the excess return is concerned, the risk-free rate rf_t is deduced from the monthly return of the portfolio. The risk-free rate is the return which could be yielded from a risk-free asset such as government T-bills (Bodie, Kane, & Marcus, 2014, p. 129).

$$\text{excess return}_{i,t} = r_{i,t} - rf_t \quad (6)$$

For each of the 36 portfolios (including the difference portfolios), excess returns for year y can be annualized by computing:

$$\text{annualized return}_{p,y} = \prod_{i=1}^{12} (1 + (r_i - r_f)) - 1 \quad (7)$$

The average monthly excess return is calculated over the entire period of 132 months (t):

$$\bar{r}_p = \frac{\sum_{t=1}^{132} (r_{p,t} - r_f)}{132} \quad (8)$$

A t-test is applied to test the hypothesis that the average monthly excess return differs significantly from 0, with the test statistic calculated as:

$$t = \frac{\bar{r}_p - \mu}{s_{\bar{r}}} \quad (9)$$

with μ=0 and the standard error $s_{\bar{r}} = \frac{SD_p}{\sqrt{n}}$.

Risk measures include volatility, which is measured by the standard deviation (SD) of stock i in year t (Bodie et al., 2014, p. 129).

$$SD_{i,t} = \sqrt{\frac{\sum (r_{i,t} - \bar{r}_i)^2}{N}} \quad (10)$$

3.2.3 Regression Analyses

After forming the difference portfolios and checking if the mean excess returns differs significantly from zero via t-test (thus, indicating a cross-sectional relationship between the sort variable and the outcome variable), time-series regression examines the pattern in excess returns. Controlling for systematic risk factors that are known to explain stock returns, does the presumable α exist for portfolios without taking on an additional risk that is inherent to these companies? A multi-factor technique (i.e. a number of specifications of it to test for robustness to various control sets) is used to investigate the relations between ESG scores and financial returns. According to the logic behind these risk factors,

the expected excess return should be higher for an asset (or portfolio) with higher exposure to one or more risk factors.

First, month-by-month regressions are performed for each portfolio p in the single factor CAPM model where the only risk factor is the market, correcting for the effect of the overall stock market return (Bali et al., 2016, p. 54) (specification I):

$$r_{p,t} - r_f = \alpha_p + \beta_p MKT_t + \varepsilon_{p,t} \quad (11)$$

The dependent variable is the excess return of portfolio p (*responsible, irresponsible, difference* with respect to the composite ESG score and subscores) in month t, α_p (the intercept) is the risk-adjusted return and expresses the over- or underperformance of the portfolio. *MKT* represents the excess market return, proxied by the „region's value-weight market portfolio" minus the risk-free rate (French, 2019). The β_p denotes the portfolio's sensitivity to the overall market return.

In specification II, the Carhart model with its size, value and momentum factor is employed. Other than the CAPM, it is built on the idea that stock returns are not only a function of the overall market movements and that cross-sectional differences are due to varying market factor sensitivities, β_p. Notably, Fama and French (1992, p. 428) emphasize that market capitalization and the book-to-market ratio add to the explanation of stock returns in the cross-section. These types of cross-sectional variation must be soaked up so as to obtain a reliable alpha expressing the over- or underperformance of the variation. The size factor, *SMB* (small minus big), acknowledges the empirical evidence according to which small stock companies tend to outperform large stock companies, and represents the returns associated with taking one unit of this size-related risk. This return is calculated from a "size factor mimicking portfolio", which is constructed by forming zero-cost value-weighted portfolios based on size (market capitalization) each year and calculating the average return on the small stock portfolios minus the average return on the big stock portfolios (Bali et al., 2016, p. 176; French, 2019). Similarly, there is strong empirical evidence that high book-to-market companies ("value stocks") outperform low book-to-market companies ("growth stocks"). The value factor, *HML*, representing the returns associated with taking on this value-related risk, is calculated as the return of a factor mimicking portfolio that is long in high book-to-market (value) stocks and short in low book-to market (growth) stocks (French, 2019). Adding again to the explanatory power of the market, size and value factor, Carhart extends the model by introducing a momentum factor which recognizes the importance of previous stock performance. The momentum factor, *MOM*, represents the return arising from a factor mimicking portfolio that is long in stocks with the highest recent performance and short in stocks with the lowest recent performance

(Bali et al., 2016, p. 54; French, 2019). Whereas the factors MKT_t, SMB_t, HML_t and MOM_t denote the return associated with the respective risk in month t (downloadable from Kenneth French's data library), the coefficients s_p, v_p, and m_p are obtained through regression analysis and denote the portfolio's sensitivity to each of these factors.

$$r_{\mathrm{p,t}} - r_f = \alpha_p + \beta_p MKT_t + s_p SMB_t + v_p HML_t + m_p MOM_t + \varepsilon_{p,t} \tag{12}$$

Specification III furthermore includes the factors for profitability and investment. While the four common risk factors of specification II are mostly used in performance studies, the profitability and investment factors are added here in the third specification to account for the "the latest insights in asset pricing theory" which helped Blitz and Fabozzi (2017) detect that there is no such thing as a premium for the "sinfulness" of assets because sin stock returns can be fully explained when these two new factors are added to asset pricing models. The profitability factor, *RMW* (robust minus weak), considers the importance of operating profitability for stock price movements, and represents the return of a factor mimicking portfolio which is long in robust operating profitability portfolios and short in weak operating profitability portfolios. Acknowledging the relevance of a firm's expected investments for its rate of return, the investment factor *CMA* (conservative minus aggressive) represents the return of a portfolio long in stocks with a conservative investment strategy and short in stock with an aggressive investment strategy (French, 2019). As before, the coefficients p_p and i_p are obtained through regression analysis and, whenever statistically significant, are used to determine whether the respective risk factor is related to the portfolio excess return.

$$r_{\mathrm{p,t}} - r_f = \alpha_p + \beta_p MKT_t + s_p SMB_t + v_p HML_t + m_p MOM_t + p_p RMW_t + i_p CMA_t + \varepsilon_{i,t} \tag{13}$$

3.2.4 Risk Analysis

The Sharpe Ratio (SR) of portfolio p is defined as its average excess return over the standard deviation of its excess returns and describes the trade-off between risk and return. When annualized it is commonly used to measure and compare the attractiveness of portfolios (Bodie et al., 2014, p. 134). It is calculated as:

$$\text{annualized SR}_\mathrm{p} = \frac{\bar{\mathrm{r}}_{\mathrm{p,t}} - \mathrm{r_f}}{\mathrm{SD}} \times \sqrt{12} \tag{14}$$

Moreover, the hypothesis is tested that socially responsible assets are associated with higher risk, measured by different moments of risk. This additional analysis accounts for the shortcoming of other studies which hardly consider higher moments of risk, as emphasized by Belghitar et al. (2014).

Since assuming "normality of excess returns hugely simplifies portfolio selection", moments of risk to be considered are skewness (third moment of risk) and kurtosis (fourth moment of risk) (Bodie et al., 2014, pp. 138–139): Assuming risk averse SRI investors, these measures of risk are particularly interesting as they account for the fact that asset (portfolio) returns are not normally distributed. Skewness (SKEW) refers to the asymmetry of the distributions around the mean and, if positive, implies that the mean is higher than the median, because the tail of the distribution extends to more positive values. Positive skewness is typically desirable as it implies that negative deviations from the mean are limited while positive ones are possible. Avoiding negative skewness is thus all the more important but standard deviation underestimates this risk of these extreme negative outliers, stressing the importance of analyzing higher moments of risk. Excess kurtosis (leptokurtic curvature) (KURT) means that the "peakedness" of the return distributions is stronger than the one of a normal distribution where KURT= 3 (mesokurtic). The resulting "fat tails" imply a higher probability of rare events, i.e. extreme positive and negative outliers or "jumps", and a lower probability of actually yielding the expected (mean) return. which is typically not preferred by investors. Skewness and kurtosis are calculated as:

$$SKEW_{i,t} = \frac{\sum\left(\frac{r_{i,t} - \overline{r_i}}{SD}\right)^3}{N} \tag{15}$$

$$KURT_{i,\mathrm{t}} = \frac{\sum\left(\frac{\mathrm{r_{i,t}} - \overline{\mathrm{r_i}}}{\mathrm{SD}}\right)^4}{\mathrm{N}} - 3 \tag{16}$$

Portfolio returns are used to calculate and compare their skewness and kurtosis and their Adjusted Sharpe Ratio (ASR): Motivated by the drawbacks of the Sharpe Ratio which assumes normally distributed returns, Pezier and White (2006) suggest an Adjusted Sharpe Ratio that considers both skewness and excess kurtosis and contains a penalty factor for the unwanted negative skewness and positive excess kurtosis in portfolio p. If the distribution is is negatively skewed and has positive excess kurtosis, the ASR gets smaller compared to the traditional SR. Annualized ASR is calculated as:

$$annualized\ ASR_p = SR_p \times \left[1 + \left(\frac{SKEW}{6}\right)SR_p - \left(\frac{KURT}{24}\right)SR_p^2\right] \times \sqrt{12} \tag{17}$$

Moreover, correlation analysis is used on a year-by-year basis to check for a potential relationship and consistency between ESG scores and subsequent stock performance in the sense of risk (volatility measured by standard deviation, kurtosis and skewness) for the entire dataset. As opposed to Clayman (2015), who uses Chi Square Frequency tests to see

if high ESG assets tend to be in high risk groups, this paper employs the Spearman rank correlation which puts emphasis on the monotony between ESG scores and risk measures. The advantage of using Spearman correlation as opposed to the Pearson correlation is that the data does not need to fulfill a normality condition and it is not very sensitive to outliers, which can otherwise easily invalidate correlation results. Conducting sporadic Jarque-Bera tests on the data set reveals that normality of the variables cannot be assumed.

As previously done with return data, ESG scores are matched with the risk measures calculated with a 6 months lag. For example an ESG score obtained in early 2010 for ESG-relevant information occurred in 2009 is matched with volatility, skewness and kurtosis obtained from return data between 06/2010-05/2011. Ranking each ESG score and each risk measure from 1 to n, the Spearman correlation coefficient takes values between -1 and 1 and is calculated as for each pair (Bali et al., 2016, p. 20):

$$\rho_t^S(X,Y) = 1 - \frac{6\ \sum_{i=1}^{n}\left(x_{i,t} - y_{i,t}\right)^2}{n_t(n_t^2 - 1)} \tag{18}$$

3.2.5 Robustness Tests

In order to test results for significance, two robustness tests are performed. In the first one, the breakpoints of the ESG portfolios are changed to the 20th and 80th percentile to decrease the range of ESG scores in a given portfolio. Return analyses as previously described are performed under these new conditions. In the second robustness tests, assets in the portfolios are equal-weighted (again based on the 30th and 70th percentile). Instead of using the respective market capitalization as the weight variable, the return of each of the n assets in a given portfolio is weighted by 1/n to calculate the portfolio return (Bali et al., 2016, p. 44). It should be noted that whereas portfolio returns are equal-weighted, the risk factors from Kenneth French's data library are not. This methodological difference must be accepted when performing the second robustness test.

3.3 Descriptive Statistics

The numbers of assets in the sample for which both Bloomberg ESG scores and corresponding return data for the respective time period could be obtained are shown in table 1. It shows that the availability does not only vary between the three regions, but also across the years: In general, availability has been increasing since inception of the rating system. Within a given year, the number of available data points mostly does not vary much between the *ESG* composite score and *ENV, SOC and GOV* subscores, suggesting that if Bloomberg adds a company to its ESG database, it rates it on all three dimensions and the

overall score. However, companies often lack an environment score when all other subscores are available, as suggested in the low number of available environment scores. Panel B shows the summary statistics for the ESG scores in the sample across all years. The average governance score is always significantly higher than the other average values, suggesting that awareness of the importance to disclose data is particularly high in this dimension. Panel C shows the Spearman correlation between the ESG variables, which is constantly positive and significant across all years. Assets with high composite ESG scores thus also have high subscores in the three dimensions, which is to be expected due to the nature of the scores. But also the subscores have high correlations with each other. It is reasonable to assume that if companies are aware of the importance of taking ESG-related measures and disclosing ESG-related information, they do so on all dimensions. As a result, the information content of the four different scores is expected to be very similar, however, since these variables are not simultaneously included in a regression model there needs to be no concern about statistical issues, only that results for portfolios built on the four different scores may be similar (Bali et al., 2016, p. 21).

The development of the average scores over the years can be seen in figure 3. It shows that in Europe, average values have almost constantly been increasing (along with available scores), but not necessarily in the other two geographies. Moreover, the extreme prominence of average governance score becomes very obvious in all three regions.

TABLE 1: ESG SUMMARY STATISTICS

Panel A reports the number of assets to be used in the sample each year. For these assets, both ESG scores and subsequent return data (and market capitalization) could be obtained. **Panel B** shows the summary statistics (mean, standard deviation, minimum and maximum value) for the ESG dataset across all years. More detailed summary statistics are provided in appendix 1. **Panel C** reports the Pearson correlation between the variables. The asterisk (*) indicates statistical significance at the 95% level.

Panel A

		2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
EU	**ESG**	123	220	251	266	312	350	365	382	392	416	422
	ENV	104	198	227	246	258	276	294	331	361	385	396
	SOC	120	218	249	262	278	298	316	347	372	403	410
	GOV	120	219	251	266	312	350	365	382	392	416	422
Japan	**ESG**	634	1094	1476	1508	1650	1725	1744	1777	1785	1784	1769
	ENV	428	781	841	876	908	933	1026	1042	1099	1118	1237
	SOC	581	935	1010	1032	1061	1082	1176	1193	1360	1507	1657
	GOV	634	1094	1475	1508	1650	1725	1744	1777	1785	1784	1769
US	**ESG**	401	1039	1308	1631	2552	2735	2852	3028	3172	3324	3376
	ENV	169	339	442	544	610	674	741	809	924	1181	1357
	SOC	361	808	1021	1151	1339	1421	1591	1724	1899	2305	2493
	GOV	400	1035	1304	1630	2551	2735	2852	3027	3172	3323	3375

Panel B

		Mean	SD	Min	Max
EU	ESG	34.24	15.96	0.83	80.58
	ENV	30.70	16.47	1.55	84.50
	SOC	38.13	17.97	3.13	82.46
	GOV	49.69	13.52	3.57	85.71
Japan	ESG	21.66	12.68	0.96	62.81
	ENV	25.23	16.06	0.78	78.29
	SOC	19.19	11.67	3.13	64.91
	GOV	43.57	7.06	3.57	67.86
US	ESG	16.88	10.01	0.83	75.62
	ENV	17.03	16.73	0.78	82.17
	SOC	15.26	12.88	3.13	86.67
	GOV	50.46	6.79	3.57	85.71

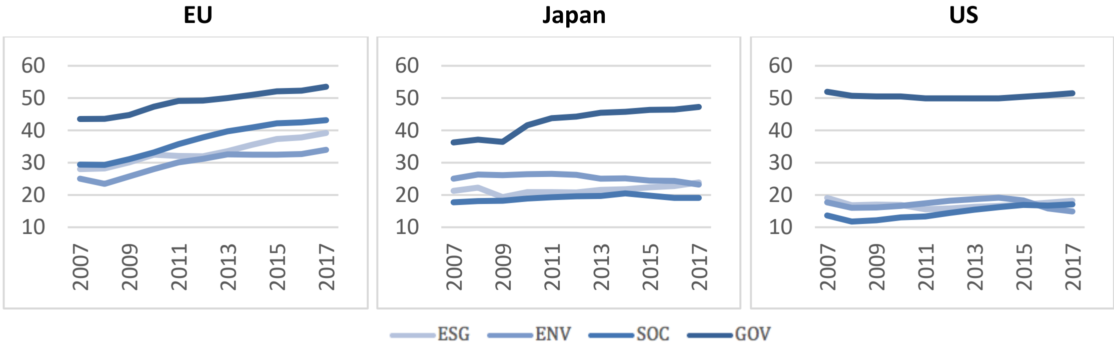
Panel C

		EU				Japan				US		
		ESG	ENV	SOC		ESG	ENV	SOC		ESG	ENV	SOC
2007	ESG	1			ESG	1			ESG	1		
	ENV	0.894*	1		ENV	0.959*	1		ENV	0.951*	1	
	SOC	0.788*	0.5317*	1	SOC	0.759*	0.660*	1	SOC	0.842*	0.683*	1
	GOV	0.581*	0.302*	0.4537*	GOV	0.455*	0.294*	0.294*	GOV	0.717*	0.580*	0.601*
2008	ESG	1			ESG	1			ESG	1		
	ENV	0.920*	1		ENV	0.959*	1		ENV	0.944*	1	
	SOC	0.850*	0.657*	1	SOC	0.753*	0.579*	1	SOC	0.982*	0.66571*	1
	GOV	0.651*	0.400*	0.536*	GOV	0.358*	0.180*	0.246*	GOV	0.686*	0.546*	0.590*
2009	ESG	1			ESG	1			ESG	1		
	ENV	0.922*	1		ENV	0.957*	1		ENV	0.943*	1	
	SOC	0.840*	0.648*	1	SOC	0.755*	0.600*	1	SOC	0.850*	0.711*	1
	GOV	0.637*	0.406*	0.486*	GOV	0.355*	0.184*	0.229*	GOV	0.711*	0.581*	0.617*
2010	ESG	1			ESG	1			ESG	1		
	ENV	0.926*	1		ENV	0.954*	1		ENV	0.928*	1	
	SOC	0.841*	0.656*	1	SOC	0.765*	0.604*	1	SOC	0.871*	0.731*	1
	GOV	0.7057*	0.510*	0.544*	GOV	0.463*	0.306*	0.353*	GOV	0.734*	0.648*	0.630*
2011	ESG	1			ESG	1			ESG	1		
	ENV	0.934*	1		ENV	0.958*	1		ENV	0.948*	1	
	SOC	0.851*	0.682*	1	SOC	0.773*	0.608*	1	SOC	0.866*	0.737*	1
	GOV	0.727*	0.550*	0.575*	GOV	0.5176*	0.364*	0.500*	GOV	0.761*	0.661*	0.680*
2012	ESG	1			ESG	1			ESG	1		
	ENV	0.931*	1		ENV	0.967*	1		ENV	0.956*	1	
	SOC	0.859*	0.697*	1	SOC	0.770*	0.616*	1	SOC	0.874*	0.763*	1
	GOV	0.741*	0.576*	0.593*	GOV	0.555*	0.432*	0.499*	GOV	0.789*	0.721*	0.667*
2013	ESG	1			ESG	1			ESG	1		
	ENV	0.943*	1		ENV	0.966*	1		ENV	0.946*	1	
	SOC	0.880*	0.756*	1	SOC	0.774*	0.615*	1	SOC	0.858*	0.732*	1
	GOV	0.720*	0.563*	0.574*	GOV	0.584*	0.470*	0.550*	GOV	0.764*	0.674*	0.637*
2014	ESG	1			ESG	1			ESG	1		
	ENV	0.942*	1		ENV	0.965*	1		ENV	0.945*	1	
	SOC	0.838*	0.700*	1	SOC	0.779*	0.622*	1	SOC	0.855*	0.733*	1
	GOV	0.696*	0.549*	0.516*	GOV	0.590*	0.480*	0.548*	GOV	0.779*	0.680*	0.682*
2015	ESG	1			ESG	1			ESG	1		
	ENV	0.942*	1		ENV	0.968*	1		ENV	0.939*	1	
	SOC	0.848*	0.703*	1	SOC	0.798*	0.654*	1	SOC	0.862*	0.743*	1
	GOV	0.699*	0.543*	0.546*	GOV	0.615*	0.517*	0.568*	GOV	0.754*	0.661*	0.650*

2016	ESG	1			ESG	1			ESG	1		
	ENV	0.939*	1		ENV	0.967*	1		ENV	0.921*	1	
	SOC	0.835*	0.687*	1	SOC	0.802*	0.662*	1	SOC	0.879*	0.747*	1
	GOV	0.695*	0.533*	0.524*	GOV	0.615*	0.522*	0.523*	GOV	0.768*	0.662*	0.657*
2017	ESG	1			ESG	1			ESG	1		
	ENV	0.939*	1		ENV	0.967*	1		ENV	0.917*	1	
	SOC	0.835*	0.682*	1	SOC	0.824*	0.694*	1	SOC	0.880*	0.740*	1
	GOV	0.671*	0.502*	0.518*	GOV	0.630*	0.544*	0.571*	GOV	0.744*	0.627*	0.626*

FIGURE 3: EVOLUTION OF AVERAGE ESG SCORES

The three graphs show the evolution of the average ESG scores and subscores for the three geographies. The y-axis represents the average scores.



The fact that environmental scores are available to a much lesser extent is also reflected in the number of assets per portfolio (based on the 30th and 70th percentile) shown in table 2. The numbers also show that in most years, the numbers of assets in the low and respective high portfolio are close and correspond with the idea that they represent the best and worst 30% of all available scores in that given year, dimension and geography. When the gap is large in some years (e.g. US *ESG* 2014) it is due to a relatively low dispersion of ESG scores in that given geography and year. If there are many companies rated with the same score on a certain dimension, they all fall into the same percentile. This is, however, a problem that commonly arises with variables that are thought to be continuous but in reality are not (Bali et al., 2016, p. 40).

TABLE 2: NUMBER OF ASSETS PER PORTFOLIO

The table presents the number of assets in each portfolio per year. The assets are assigned to the portfolios based on their ESG scores. If they belong, in a given ESG dimension, to the bottom (top) 30%, they are assigned to the irresponsible (responsible) portfolios.

			2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
EU	**ESG**	**irresp.**	37	66	75	79	92	104	109	114	117	123	123
		respon.	37	66	73	77	93	104	108	113	117	124	127
	ENV	**irresp.**	31	55	66	74	77	83	88	98	108	114	113
		respon.	31	60	67	74	75	79	85	99	104	114	118
	SOC	**irresp.**	36	63	74	78	83	88	70	99	102	110	121
		respon.	32	65	71	76	80	89	92	85	110	116	119
	GOV	**irresp.**	36	64	68	79	87	100	92	113	106	113	119
		respon.	24	55	71	58	85	88	104	110	110	124	79
Japan	**ESG**	**irresp.**	173	321	435	276	210	185	499	508	528	487	506
		respon.	187	321	430	444	491	516	522	530	536	528	529
	ENV	**irresp.**	128	233	249	255	270	257	308	313	306	315	370
		respon.	127	222	242	258	272	265	294	303	326	325	365
	SOC	**irresp.**	82	144	159	154	150	321	341	317	399	249	264
		respon.	167	274	303	307	207	218	233	274	295	321	495
	GOV	**irresp.**	136	301	278	414	365	302	142	116	522	491	221
		respon.	177	320	411	204	275	282	315	331	342	338	367
US	**ESG**	**irresp.**	69	182	228	279	110	115	142	173	125	527	569
		respon.	75	186	216	275	408	474	500	540	601	619	619
	ENV	**irresp.**	30	52	76	92	103	105	120	144	153	165	214
		respon.	30	58	76	93	105	118	130	142	167	220	253
	SOC	**irresp.**	66	44	51	54	220	227	269	265	253	315	306
		respon.	63	122	168	197	201	247	195	243	282	333	372
	GOV	**irresp.**	61	72	105	101	153	161	191	229	196	216	207
		Respon.	71	113	145	173	204	229	247	275	333	419	500

4. Empirical Results

4.1 Return Analysis

For a preliminary overview of the return characteristics of the portfolios, annualized excess returns of the 36 portfolios (long in irresponsible stocks, long in responsible stocks, long-short strategy) are provided in figure 4. A more extensive version is provided in appendix 2. In line with the correlation analysis revealing a strong association between the four different ESG scores, the graphs support the suspicion that the results do not differ greatly between the portfolios in a given geography. Indeed, the annualized excess return curve looks very similar in each of the sustainability dimensions. Secondly, the graphs for the responsible and irresponsible portfolios often move in sync along general trends in equity markets (e.g. the US bear market between late 2007 and early 2009), suggesting that general market movements have a stronger influence on the portfolio performance than their level of responsibility. At first sight the irresponsible portfolios seem to slightly outperform in Japan and the US (in Europe, outperformance by the low portfolios only appears in the first half of the observation period), as also indicated by the performance of the difference portfolio that goes long in irresponsible and short in responsible portfolios. These observations lend preliminary support to the hypothesis that there is a financial penalty for investing responsibly in these investment segments.

Next, the time-series averages and respective t-statistics are analyzed for further indication of a relationship between irresponsible and responsible assets and returns, because they represent estimations of the true average values of the outcome variable for assets in each of the portfolios in the average time period of 132 months. For the responsible and irresponsible portfolios, the time-series averages represent the mean returns of the strategy that goes long in the portfolio's respective assets. The focus lies on the time-series mean of the difference portfolio, because it serves as an estimate of the difference of the average value of the returns for assets in the responsible portfolio compared to those in the irresponsible portfolio. T-statics are used to test the hypothesis that the average time-series mean for the difference portfolios is equal to zero; if this hypothesis can be rejected at the 95% significance level, there is evidence of a cross-sectional relationship between the sorting variables (ESG scores) and the outcome variable (excess returns), which would support the hypothesis of financial penalty for high-ESG assets.

Mean excess returns of the 132 months in the observation period for the 12 difference portfolios are reported in table 3. Only in a few cases, that is for the *ENV* difference portfolio in Japan, the *ESG* and *GOV* difference portfolio sin the US the t-test is able to reject the hypothesis that the time-series mean excess return differs significantly from 0. In all other cases, the approach fails to detect this statistically positive or negative returns

and a cross-sectional relation between ESG scores and excess returns. Time-series averages of the long-only portfolios are reported in appendix 3. Without going into the details, it can be seen that almost all long-only portfolios have on average a significantly positive monthly excess return in the US and in Japan.

FIGURE 4: ANNUALIZED EXCESS RETURNS OF THE 36 PORTFOLIOS

The figures present the annualized excess returns of all 36 portfolios (irresponsible, responsible, difference in 12 investment segments). The x-axis reports the years, i.e. the reference year when the portfolios were built based on available ESG scores. The y-axis denotes the annualized excess return, for each geography and the four different ESG scores. Note the different scaling of the y-axis, arising from different ranges of annualized excess returns.

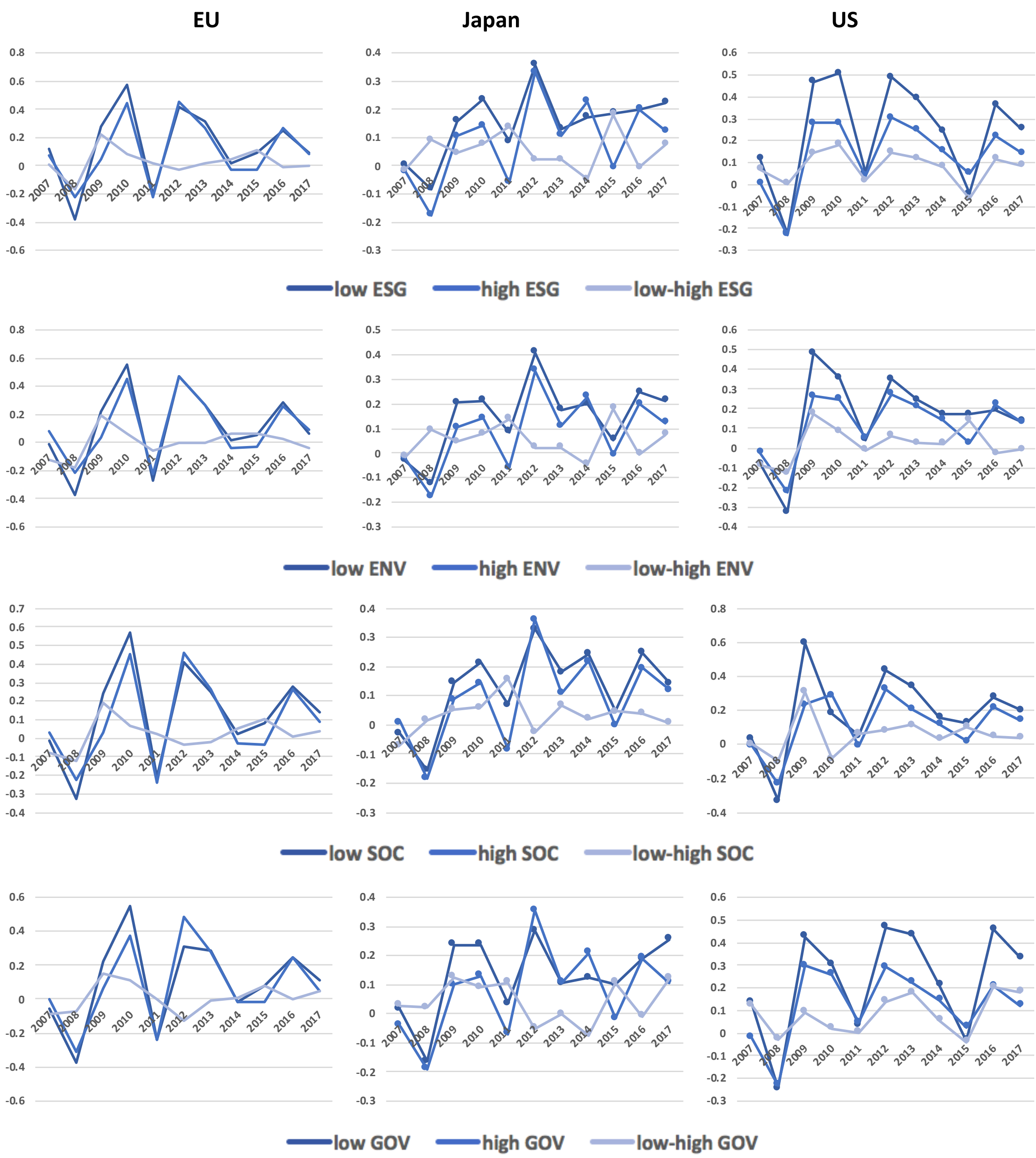


TABLE 3: AVERAGE MONTHLY EXCESS RETURNS AND T-TESTS OF THE 12 DIFFERENCE PORTFOLIOS

The table reports the time-series averages for the 132 monthly excess returns of the 12 difference portfolios. T-tests are performed and reported to test the hypothesis that the mean excess returns differ statistically from zero, indicating a cross-sectional relation between ESG scores and excess returns. The asterisk (*) indicates statistical significance at the 95% level (t-statistics > 2.00).

		av. monthly excess return	t-test
EU	**ESG diff.**	0.22%	0.8972
	ENV diff.	-0.01%	-0.0290
	SOC diff.	0.17%	0.7626
	GOV diff.	0.07%	0.3522
Japan	**ESG diff.**	0.44%	1.721
	ENV diff.	0.39%*	2.1057
	SOC diff.	0.28%	1.4489
	GOV diff.	0.34%	1.6432
US	**ESG diff.**	0.66%*	2.5426
	ENV diff.	0.17%	0.7883
	SOC diff.	0.48%	1.3359
	GOV diff.	0.71%*	2.4252

Finally, the regression outputs are considered to get more insight into the cross-sectional relationship between ESG scores and excess portfolio returns. The focus in this section is on the alpha of the difference portfolio: It expresses the return difference between under- and overperformers with respect to a particular ESG score and is used to see if a strategy that goes long in irresponsible (low scores) and short in responsible (high scores) assets can yield a positive alpha, which would support the hypothesis that there is financial penalty for investing responsibly. Results of the model specifications (I)-(III) for all 12 investment segments are reported in table 4. First, it should be noted that the explanatory power, as measured by the adjusted R^2, increases with the model specifications in almost all investment segments (except *GOV* Europe), which was to be expected as the additional risk factors have proven in previous research to explain stock returns. For example, only 4,4% of the cross-section in stock returns of the *ESG* difference portfolio in Japan can be explained by the CAPM, where the market factor is the only risk factor. Adding the size, value, momentum, investment and profitability factors, explanatory power rises to 55,7%. Interestingly, in all of the US portfolios, the explanatory power remains below 20% in all model specifications, indicating a relatively broad scatter around the trend line.

Turning now to the performance measurements, the intercept α_I is interpreted as the average excess return of the portfolio that cannot be explained by the portfolio's exposure to any of the risk factors that are included in the respective model. The following interpretations for α_i hold:

- α_i <0: the return was too low for the risk taken
- α_i =0: the return was adequate for the risk taken
- α_i >0: excess return has been achieved for the risk taken

The first column reports the results of the CAPM (specification I) for each geography, ESG composite score and subscores. In Europe, none of the difference portfolios can yield a significantly positive alpha in the CAPM. This is in line with the preliminary results from the annualized returns and the insignificant t-statistics, both of which suggested that there is no performance difference between the portfolios and thus, an insignificant performance by the difference portfolios. In Japan, the CAPM shows a significantly positive alpha in the *ESG* and *ENV* difference portfolios, in the US, the alpha is significantly positive in the *ESG* and *GOV* dimension. These results have been mostly predicted by the t-tests of the average monthly excess returns.

In order to determine whether the outperformance of the difference portfolio is persistent, further risk factors are added. Results are presented in model specification (II) and (III). In Japan, the alphas disappear in both more sophisticated model specifications. If coefficients are statistically significant in one specification but become insignificant when additional control factors are added to the model, then the relation between the dependent variable (excess returns) and the explanatory variables can be explained by some linear combination of the added control variables (Bali et al., 2016, p. 97). Indeed, various added control variables show significant loadings in all specifications, most importantly the size factor. In the US, the significant alpha remains in the *ESG* and the *GOV* difference portfolio, even controlling for various risk factors. While the R^2 remains very small, i.e. there is a broad variability around the main trend, the alphas even grown in comparison with the CAPM, suggesting that these additional risk factors need to be controlled for in other to detect the true relation of interest. Investing in the US portfolio that contains the assets with the lowest 30% of all *ESG* composite scores (*GOV* scores) in each year and shorting the portfolio with the highest 30% of *ESG* composite scores (*GOV* scores) would have yielded an excess return of 0.64% (0.8%) per month, which corresponds to a yearly overperformance of almost 8% (10%) in these two investment segments. While this effect is impressive, it also remains the only evidence in this study that might experience a return-related penalty when investing responsibly: When investing in high-*ESG* and high-*GOV* assets in the US. It follows from the regression analaysis of the difference portfolios that the notion of “compensation effects” seems widely misleading at this point, in the sense that in almost all investment segments in this study there is no extra-return to be gained from the strategy of shorting responsible stocks and buying irresponsible stocks, and most responsible investors can equally enjoy the feeling of holding high ESG scores and

comparably returns than low-ESG assets, without having to be "compensated" by the non-financial utility gained from the responsible nature of their assets.

TABLE 4: REGRESSION RESULTS FOR THE DIFFERENCE PORTFOLIOS

The table presents the risk-adjusted alphas (intercept), factor sensitivities for the risk factors size (SMB), value (HML), momentum (MOM), profitability (RMW) and investment (CMA), and the adjusted R^2 for the 12 zero-cost difference portfolios which are long in the irresponsible and short in the responsible assets based on the $30^{th}/70^{th}$ percentiles. For each investment segment, outputs of the three model specifications (I)-(III) are reported. The asterisk (*) indicates statistical significance at the 95% level.

		EU			Japan			US		
		(I) CAPM	(II) Carhart	(III) 6 Factor	(I) CAPM	(II) Carhart	(III) 6 Factor	(I) CAPM	(II) Carhart	(III) 6 Factor
ESG	**α**	0.19932	0.13218	0.19577	0.49672*	0.12679	0.1941	0.53628*	0.54134*	0.63819*
diff.	**MKT-Rf**	0.07462	0.09036*	0.04651	-0.15839*	-0.07630	-0.09262*	0.18340*	0.07711	0.04719
	SMB		0.91358*	0.85208*		0.73465*	0.71413*		0.41519*	0.36258*
	HML		-0.19581	-0.05186		-0.12567	-0.20632*		-0.11043	-0.17110
	MOM		-0.13135*	-0.08894		0.18434*	0.20881*		-0.10444	-0.10772
	RMW			-0.02746			-0.29916*			-0.29732
	CMA			-0.35356			-0.06870			0.1554
	Adj. R2	*0.01523*	*0.4118*	*0.419*	*0.04409*	*0.5482*	*0.5573*	*0.06443*	*0.1646*	*0.1724*
ENV	**α**	-0.04567	-0.04846	0.04179	0.38005*	0.21196	0.27747	0.09666	0.12756	0.2284
diff.	**MKT-Rf**	0.13072*	0.10087*	0.07183	0.03971	0.06326	0.04758	0.10790*	0.06205	0.02843
	SMB		0.88561*	0.83384*		0.35450*	0.33442*		0.06862	0.0193
	HML		-0.03201	-0.01403		-0.16951*	-0.25174*		0.00438	-0.03489
	MOM		-0.15862*	-0.13149*		0.03182	0.05449		-0.07703	-0.07810
	RMW			-0.15092			-0.29585*			-0.29242
	CMA			-0.23083			-0.06120			0.09245
	Adj. R2	*0.06573*	*0.4704*	*0.4706*	*0.013*	*0.2165*	*0.2341*	*0.02823*	*0.03242*	*0.04448*
SOC	**α**	0.17229	0.03734	0.135973	0.280501	0.0827	0.13996	0.28641	0.40911	0.61032
diff.	**MKT-Rf**	-0.01162	0.03893	0.006383	-0.005935	0.02652	0.13996	0.28552*	0.16228	0.0993
	SMB		0.87878*	0.821328*		0.40583*	0.13996*		0.07813	-0.02938
	HML		-0.24784*	-0.223593		-0.08348	-0.18339*		0.15066	0.03247
	MOM		-0.05972	-0.029247		0.02378	0.03499		-0.19415*	-0.20019*
	RMW			-0.162163			-0.29396*			-0.61208*
	CMA			-0.258945			-0.01117			0.30009
	Adj. R2	*-0.007001*	*0.4403*	*0.4441*	*-0.00756*	*0.1908*	*0.2092*	*0.08283*	*0.1385*	*0.1706*
GOV	**α**	0.06289	-0.02605	-0.03676	0.35498	0.177823	0.13659	0.6196*	0.58855*	0.80176*
diff.	**MKT-Rf**	0.01933	0.04971	0.04206	-0.04119	0.003858	0.13659	0.1288	0.08186	-0.00630
	SMB		0.64202*	0.63629*		0.351134*	0.36073*		0.24547	0.17921
	HML		-0.15989	-0.09973		-0.159033*	-0.20462*		-0.20450	-0.12109
	MOM		-0.04933	-0.04150		0.169532*	0.12538*		-0.09031	-0.07621
	RMW			0.05533			0.06349			-0.49889*
	CMA			-0.06320			0.18555			-0.28749
	Adj. R2	*-0.005223*	*0.2922*	*0.2825*	*-0.00234*	*0.2952*	*0.2966*	*0.02066*	*0.04309*	*0.08989*

In view of the widely insignificant alphas of the difference portfolios and the interpretation of these findings, a look at the results of the long-only portfolios seems worthwhile (see table 5). As indicated by the R^2, the explanatory power is very strong in all models. At least 65%, but mostly a much higher percentage of the variation in stock returns can be explained by the models. The improvement in explanatory power is strong between the CAPM and the Carhart model, however, less strong between the Carhart model and the 6-factor model. The results also show that both responsible and irresponsible portfolios, in 11 out of 12 investment segments, yield constantly significantly positive alphas in the CAPM; when adding size-, value-, momentum- and profitability-related risk factors, alphas are significantly positive in all 12 investment segments, and the explanatory power is high. Also, in all cases, the alphas resulting in the 6-factor model are higher in the irresponsible portfolios than in the responsible portfolios but only in the US *ENS* and *GOV* dimension, the difference is high enough to yield a significantly positive alpha in the difference portfolio. In all other investment segments, the insignificance in the alphas of the difference portfolios suggests that the hypothesis of an insignificant relationship between Bloomberg's ESG scores and superior returns cannot be rejected, and that it is hardly possible to detect a "compensation effect" for high ESG assets.

With these two findings in mind, it seems more reasonable to assume that there is a strong link between a company's likelihood to have ESG data disclosed by Bloomberg and their financial result. As stated previously, the link between these two factors might reasonably be firm size. Looking at the factor loadings on the SMB factor, it turns out that the responsible portfolios are weighted towards large-cap stocks (as indicated by the negative sign), whereas most irresponsible portfolios are weighted towards small-cap stocks. Since alphas are significantly positive in responsible and irresponsible portfolios, the size factor does not allow for a stringent explanation attempt. Another possible explanation is that Bloomberg tends to report ESG-relevant information on historically financially successful firms[6] which investors might be more interesting in obtaining non-financial information, or even that these firms have more resources to publish ESG-relevant information that Bloomberg can use to calculate the disclosure score. As will be discussed in more detail in chapter 5, it is important to keep in mind that the performance analysis in this study employs a sample of assets that are rated with ESG scores, and uses risk factors that are built on the overall market. Regarding the other risk factors, there are no significant patterns to be seen across all investment segments. The only factor that is consistently significantly positive in the long-only portfolios is exposure to systematic market risk, as measured by beta, which is slightly higher in most of the irresponsible portfolios, but the

[6] If this hypothesis is true and some of the assets in the portfolios have a very large market capitalization, then their financial outperformance tends to be overstated in the value-weighted portfolios used in this study. These effects are limited by using equal-weighted portfolios in the robustness tests, see chapter 4.3.

difference is not significant. It remains the finding that both high- and low ESG assets are high-beta stocks.

TABLE 5: REGRESSION RESULTS FOR THE LONG-ONLY PORTFOLIOS

The table presents the risk-adjusted alphas (intercept), factor sensitivities for the risk factors size (SMB), value (HML), momentum (MOM), profitability (RMW) and investment (CMA), and the adjusted R^2 for the 24 long-only portfolios based on the $30^{th}/70^{th}$ percentiles. For each investment segment, outputs of the three model specifications (I)-(III) are reported. **Panel A** reports results for the responsible portfolios, **panel B** reports results for the irresponsible portfolios. The asterisk (*) indicates statistical significance at the 95% level.

Panel A

		EU			Japan			US		
		(I) CAPM	(II) Carhart	(III) 6 Factor	(I) CAPM	(II) Carhart	(III) 6 Factor	(I) CAPM	(II) Carhart	(III) 6 Factor
ESG	α	0.53389*	0.629307*	0.54035*	0.42139*	0.49634*	0.46059*	0.41553*	0.43226*	0.42574*
resp.	MKT-Rf	1.01406*	0.983843*	1.00461*	0.96099*	0.94737*	0.95843*	0.99618*	1.01092*	1.01231*
	SMB		-0.521808*	-0.47918*		-0.15510*	-0.14559*		-0.13038*	-0.12544*
	HML		0.102786	0.12911		0.08161*	0.08003*		0.03864	0.04885
	MOM		0.005878	-0.01305		-0.03889*	-0.06552*		-0.01195	-0.01112
	RMW			0.1752			0.10290*			0.0244
	CMA			0.16345			0.10358*			-0.02824
	Adj. R2	*0.9216*	*0.9475*	*0.948*	*0.9573*	*0.9723*	*0.9735*	*0.9708*	*0.9747*	*0.9744*
ENV	α	0.54839*	0.63369*	0.54130*	0.46471*	0.52180*	0.44211*	0.32858*	0.35877*	0.349548*
resp.	MKT-Rf	0.99842*	0.97108*	0.99196*	0.91188*	0.91054*	0.93556*	0.94373*	0.96267*	0.963875*
	SMB		-0.55827*	-0.51473*		--0.12929*	-0.10832*		-0.19016*	-0.181493*
	HML		0.12222	0.15345		0.09116*	0.08084		0.08108*	0.102874*
	MOM		0.02916	0.01019		0.02289	-0.03856		-0.01169	-0.009807
	RMW			0.18432			0.22079*			0.039807
	CMA			0.16411			0.24113*			-0.061289
	Adj. R2	*0.9063*	*0.937*	*0.9374*	*0.9249*	*0.933*	*0.9409*	*0.9404*	*0.9499*	*0.9495*
SOC	α	0.48878*	0.60668*	0.4852*	0.41559*	0.50626*	0.49762*	0.30745*	0.337082*	0.315602*
resp.	MKT-Rf	1.04090*	0.99921*	1.02003*	0.97720*	0.95546*	0.95832*	0.95681*	0.969288*	0.978263*
	SMB		-0.53659*	-0.48645*		-0.17956*	-0.17738*		-0.141043*	-0.134574*
	HML		0.12746	0.2058		0.05326*	0.04938		0.102538*	0.093231*
	MOM		-0.01135	-0.02968		-0.06593*	-0.07344*		0.008216	0.006707
	RMW			0.26473			0.02047			0.049613
	CMA			0.16156			0.03031			0.031586
	Adj. R2	*0.9183*	*0.9447*	*0.9458*	*0.9537*	*0.9728*	*0.9725*	*0.9528*	*0.9588*	*0.9585*
GOV	α	0.35969*	0.48436*	0.454442*	0.34561*	0.43019*	0.431641*	0.34132*	0.38156*	0.37065*
resp.	MKT-Rf	1.07333*	1.03053*	1.056703*	0.96522*	0.94614*	0.945461*	0.96902*	0.97524*	0.98059*
	SMB		-0.36331*	-0.328454*		-0.17139*	-0.171649*		-0.16004*	-0.15853*
	HML		0.06308	-0.035727		0.09784*	0.102097*		0.10919*	0.09670*
	MOM		-0.06626*	-0.091731*		-0.07265*	-0.070277*		-0.0141	-0.01563
	RMW			-0.005754			0.001095			0.01961
	CMA			0.211538*			-0.010548			0.03857
	Adj. R2	*0.9529*	*0.9654*	*0.966*	*0.9368*	*0.9602*	*0.9596*	*0.9546*	*0.963*	*0.9626*

Panel B

		EU			Japan			US		
		(I) CAPM	(II) Carhart	(III) 6 Factor	(I) CAPM	(II) Carhart	(III) 6 Factor	(I) CAPM	(II) Carhart	(III) 6 Factor
ESG	**α**	0.77901*	0.80886*	0.78625*	0.96464*	0.67222*	0.70613*	0.99952*	1.02047*	1.11122*
irresp.	**MKT-Rf**	1.08673*	1.07148*	1.04723*	0.79896*	0.86732*	0.86162*	1.17601*	1.08501*	1.05593*
	SMB		0.38158*	0.36084*		0.57400*	0.56219*		0.28623*	0.23926*
	HML		-0.08872	0.08366		-0.04348	-0.13095		0.07205	-0.11874
	MOM		-0.12441*	-0.09982		0.14759*	0.14555*		0.1144	-0.11648*
	RMW			0.14408			-0.20978			-0.2713
	CMA			-0.19943			0.03612			0.11614
	Adj. R2	*0.9063*	*0.9217*	*0.9221*	*0.6525*	*0.797*	*0.7994*	*0.7612*	*0.7756*	*0.7765*
ENV	**α**	0.54852*	0.63260*	0.63322*	0.89129*	0.78285*	0.77102*	0.472944*	0.5332*	0.62523*
irresp.	**MKT-Rf**	1.12719*	1.06922*	1.05990*	0.94794*	0.97005*	0.97896*	1.048052*	1.02169*	0.98874*
	SMB		0.31716*	0.30705*		0.21966*	0.21975*		-0.12013	-0.16008
	HML		0.09451	0.14584		-0.07776	-0.17555*		0.0852	0.0715
	MOM		-0.12840*	-0.11913*		0.05686	0.01819		0.08674	-0.08554
	RMW			0.02973			-0.08858			-0.25099
	CMA			-0.07605			0.18117			0.02014
	Adj. R2	*0.9189*	*0.9331*	*0.9322*	*0.8215*	*0.8419*	*0.8456*	*0.7967*	*0.8033*	*0.8049*
SOC	**α**	0.70688*	0.69139*	0.67131*	0.74262*	0.63805*	0.68902*	0.64157	0.79306*	0.97321*
irresp.	**MKT-Rf**	1.02733*	1.03542*	1.02252*	0.96763*	0.97824*	0.96846*	1.23875*	1.12853*	1.07399*
	SMB		0.33200*	0.32281*		0.22071*	0.20366*		-0.0615	-0.16184
	HML		-0.11608	-0.01138		-0.02963	-0.13866		0.25295	0.12921
	MOM		-0.07001	-0.05676		-0.04001	-0.03619		-0.18395*	-0.19112*
	RMW			0.0989			-0.28701*			-0.56084*
	CMA			-0.1067			0.02038			0.32065
	Adj. R2	*0.9225*	*0.9339*	*0.9335*	*0.8364*	*0.8452*	*0.8509*	*0.6615*	*0.6884*	*0.6991*
GOV	**α**	0.46839*	0.50568*	0.46781*	0.7471*	0.65709*	0.61967*	1.00857*	1.01697*	1.21969*
irresp.	**MKT-Rf**	1.09072*	1.07752*	1.09487*	0.9204*	0.94625*	0.96025*	1.09427*	1.05408*	0.97072*
	SMB		0.26852*	0.29577*		0.17419*	0.18273*		0.08685	0.0228
	HML		-0.09251	-0.12904		-0.06061	-0.10718		-0.09556	-0.02087
	MOM		-0.11453*	-0.13106*		0.09903*	0.05736		-0.10243	-0.08948
	RMW			0.04591			0.05106			-0.47765*
	CMA			0.13902			0.17624			-0.25994
	Adj. R2	*0.8973*	*0.905*	*0.904*	*0.8006*	*0.8211*	*0.8218*	*0.6723*	*0.6714*	*0.6846*

4.2 Risk Analysis

Table 6 reports moments of risk and the Sharpe Ratios for the responsible and irresponsible portfolios in all 12 investment segments.

Standard deviation shows that volatility is mostly higher in the irresponsible portfolios, especially in the US, to a lesser degree in Europe and not as remarkably in Japan. The annualized Sharpe Ratio, based on the portfolios' volatilities over the 132 months, is higher

for the irresponsible portfolios in 10 out of 12 segments. The effect of higher time-series monthly average returns in the irresponsible portfolios[7] (refer back to 4.1) seems to outweigh the effect of slightly higher standard deviation, making irresponsible portfolios overall more attractive that their responsible counterparts by this measure.

Moving on to skewness in the 132 monthly portfolio returns, it can be seen that all responsible portfolios have faintly negative skewness (-0.5>SKEW>0), which is typically considered to be still approximately symmetric. The exception is the *SOC* dimension in the US, where the responsible portfolio is moderately negatively skewed (-1<SKEW<-0.5): This pronunciation of the left tail suggests that deviations from the mean in the responsible portfolio tend to be slightly negative. In practical terms, US investors holding the *SOC* responsible portfolios are rather confronted with negative than with positive monthly return surprises, which is clearly not preferred. In Europe, all four irresponsible portfolios show a stronger negative skewness than responsible portfolios, and in the *ESG*, *SOC*, and *GOV* dimension, the negative skewness is considered moderate. In that sense, investing in the responsible assets instead of the irresponsible equivalents can be seen as a limited protection against negative deviations from the mean in Europe. In Japan, also the irresponsible portfolios are only very faintly skewed and thus still considered approximately symmetric. In the US, the irresponsible portfolios are always faintly positively skewed. Only in the *SOC* dimension, the positive skewness can even be considered strong (SKEW>1); therefore, investing in the irresponsible *SOC* portfolio in the US can be seen as a protection against the negative deviations in the responsible equivalent, while the distribution also promises some positive surprises.

Excess kurtosis in Europe's irresponsible portfolio is very strong (KURT>1), suggesting a strong peakedness, whereas the responsible portfolios are close to mesokurtic. Undesirable rare events seem indeed more likely in the irresponsible portfolios; together with the findings on skewness, this supports the idea that responsible assets are more attractive with respect to their higher moments of risk, at least in Europe (and especially the *ESG*, *SOC* and *GOV*) dimension. In Japan, again, the distributions are close to mesokurtic with the only exception of the irresponsible *ESG* portfolio, showing fat tails. In the US, excess kurtosis is large in most portfolios but the peakedness tends to be stronger in the irresponsible portfolios, with an especially extreme value in the *SOC* dimension: Together with the skewness of the distribution, the result stems from some extremely positive outliers in the irresponsible portfolios. In the *GOV* dimension, excess kurtosis is stronger in the responsible portfolio.

[7] Recall that time-series monthly averages were not statistically different from 0 in Europe.

Including the penalty factors for negative skewness and positive excess kurtosis in the annualized Adjusted Sharpe Ratio, the judgement about the attractiveness of portfolios does not change much in comparison with the Sharpe Ratio. Even if the values slightly change for the portfolios that show moderate or strong skewness and kurtosis, the irresponsible portfolios still dominate their responsible counterparts in all investment segments except EU *ENV*. As Japan's portfolios all have an approximately normal distribution, there is no real difference between the Sharpe Ratio and the Adjusted Sharpe Ratio.

In sum, the analysis of moments of risk on portfolio level suggests that in Europe, investing responsible brings utility in the sense of lower volatility and that extreme and negative events tend to be fewer in the *ESG*, *SOC* and *GOV* dimensions. This finding makes sense in combination with certain risks (e.g. litigation risk) that could be associated with the level of responsibility of assets, but there is no evidence of "compensation effects". In the US, volatility is lower among responsible portfolios and investments in irresponsible assets tend to be riskier as measured by excess kurtosis (except in the *GOV* dimension). However, the irresponsible *SOC* dimension tends to have positive deviations from the mean as measured by positive skewness. In Japan, the portfolios are close to normally distributed and there is no statement to be made about loss in utility in higher moments of risk or volatility. Put into relation with average excess returns, the relatively weak findings on skewness and kurtosis also in the US and Europe are put into perspective, as the Adjusted Sharpe Ratio still suggests that irresponsible portfolios are in general more attractive despite the slight tendencies for negative deviations and fat tails. Again, this finding loses some of its relevance when considering the general interpretations of Sharpe Ratios. Only very few (Adjusted) Sharpe Ratios are larger than 1 and are thus considered "acceptable": Japan *ESG* irresponsible, and US *ESG* and *GOV* irresponsible. Nonetheless, only considering the Adjusted Sharpe Ratio, responsible investors indeed accept losing utility from a bad ratio compared to the one they could get from investing in the irresponsible portfolios.

Next, moments of risk are analyzed on stock-level via correlation analysis. The results of the Spearman rank correlation analysis between ESG scores and higher risk moments are reported in table 7. They are split into years so as to detect a possible change in the relationship over the years.

Overall, evidence of significant correlations between ESG scores and measures of risk is very low. In Europe, the only observation is that in later years in the observation period there seems to be a slightly negative correlation between *ESG* and *ENV* scores and volatility which supports the finding from the portfolio analysis. However, the preliminary findings

of a less negative skewness and lower kurtosis cannot be supported on stock-level, due to very sporadic significant results and no clear trend in the signs.

In Japan, the correlation analysis on stock-level finds a weak, but significantly negative association between *ESG* and *ENV* scores and volatility. Moreover, in all dimensions there is a weak but significantly negative correlation between the level of responsibility of an asset and kurtosis and, in later years, also skewness. Thus, there is weak evidence that extreme return events are rare in assets with high scores, but also that the deviations tend to be negative. However, significance levels are low in half of the years to make a statement at all. In the US, there is evidence for a negative relation between volatility and responsible assets as shown by a great number of significantly negative coefficients, and it supports the findings on portfolio level. However, kurtosis seems to be completely unrelated to ESG scores, and there is only light evidence that skewness might be negatively related to it.

TABLE 6: RISK MOMENTS AND (ADJUSTED) SHARPE RATIOS FOR THE LONG-ONLY PORTFOLIOS

The table presents the standard deviation (SD), skewness (Skew) and kurtosis (Kurt) for the 24 long-only portfolios. Moreover, the annualized Sharpe Ratio (SR) and Adjusted Sharpe Ratio (ASR) are reported.

			SD	SR (annualized)	Skew	Kurt	ASR (annualized)
EU	**ESG**	**irresp.**	0.06514	0.58523	-0.50863	2.30103	0.57525
		resp.	0.06028	0.47917	-0.11228	-0.04978	0.47795
	ENV	**irresp.**	0.06711	0.45528	-0.45961	2.10661	0.45001
		resp.	0.05985	0.48838	-0.11610	-0.07108	0.48707
	SOC	**irresp.**	0.06104	0.57361	-0.61136	1.78777	0.56276
		resp.	0.06199	0.44522	-0.08138	-0.07820	0.44446
	GOV	**irresp.**	0.06571	0.41705	-0.54458	2.09602	0.41197
		resp.	0.06276	0.37380	-0.34883	0.73238	0.37133
Japan	**ESG**	**irresp.**	0.04136	1.04694	-0.17033	2.51956	1.02792
		resp.	0.04114	0.64367	-0.24748	0.80807	0.63799
	ENV	**irresp.**	0.04378	0.97298	-0.07340	0.60511	0.96770
		resp.	0.03971	0.68934	-0.32442	0.84758	0.68096
	SOC	**irresp.**	0.04430	0.85094	0.02140	0.49924	0.85062
		resp.	0.04192	0.63180	-0.24917	0.94329	0.62618
	GOV	**irresp.**	0.04306	0.86546	-0.23462	1.19252	0.85432
		resp.	0.04177	0.57242	-0.20667	0.64772	0.56874
US	**ESG**	**irresp.**	0.05864	1.07193	0.18471	2.54004	1.07128
		resp.	0.04404	0.87006	-0.47704	1.43918	0.84939
	ENV	**irresp.**	0.05110	0.81311	0.03375	1.54411	0.81130
		resp.	0.04238	0.80326	-0.42979	1.11409	0.78791
	SOC	**irresp.**	0.06622	0.78482	1.25264	9.48167	0.80602
		resp.	0.04269	0.78764	-0.56479	1.59453	0.76808
	GOV	**irresp.**	0.05803	1.05489	0.10977	0.65078	1.05811
		resp.	0.04320	0.81240	-0.47316	1.20083	0.79514

TABLE 7: SPEARMAN CORRELATION BETWEEN ESG SCORES AND RISK MOMENTS

Panel A reports the Spearman correlation coefficients between ESG scores and Standard Deviation, **panel B** reports the Spearman correlation coefficients between ESG scores and skewness, **panel C** reports the Spearman correlation coefficients between ESG scores and kurtosis. The asterisk (*) indicates statistical significance at the 95% level.

Panel A

	EU				Japan				US			
	ESG	ENV	SOC	GOV	ESG	ENV	SOC	GOV	ESG	ENV	SOC	GOV
2007	-0.0724	-0.0303	-0.1213	-0.0539	-0.0143	0.0009	-0.0528	-0.0233	-0.1899	-0.1697	-0.1903	-0.1551
2008	-0.0248	-0.0168	-0.0115	-0.068	0.0517	0.0469	0.0597	0.008	-0.1484*	-0.1378	-0.1775*	-0.0529
2009	-0.0395	-0.036	0.0228	-0.1117	-0.0863*	-0.0606	-0.0885*	-0.0658	-0.1360*	-0.1528*	-0.1358*	-0.0818
2010	0.0325	0.0336	0.036	0.0518	-0.1041*	-0.0801*	-0.1328*	-0.0534	-0.0755	-0.1018	-0.0343	-0.0704
2011	0.029	0.055	0.0335	-0.0217	0.0057	0.0212	-0.0394	0.0186	-0.1696*	-0.1599*	-0.1168*	-0.1918*
2012	-0.0444	-0.0282	-0.0515	0.0102	-0.0008	0.0041	-0.0224	0.0604	-0.1422*	-0.1494*	-0.0852	-0.2095*
2013	0.0079	-0.0092	0.0473	0.0489	-0.0718*	-0.0679*	-0.0465	-0.0593	-0.1787*	-0.1695*	-0.1389*	-0.2123*
2014	-0.1155*	-0.1168*	-0.0607	-0.0679	-0.0296	-0.0325	-0.0046	-0.0388	-0.1872*	-0.1676*	-0.1047*	-0.2111*
2015	-0.0448	-0.0219	-0.0243	-0.086	0.0792*	0.0786*	0.0726*	0.0863*	-0.1861*	-0.1843*	-0.1221*	-0.2112*
2016	-0.1369*	-0.1633*	-0.0526	-0.0383	0.0175	0.006	0.0427	0.0006	-0.2165*	-0.2267*	-0.1877*	-0.2558*
2017	-0.1555*	-0.1554*	-0.0533	-0.1934*	-0.0575*	-0.0686*	-0.0141	-0.0893*	-0.1762*	-0.1958*	-0.1271*	-0.1775*

Panel B

	ESG	ENV	SOC	GOV	ESG	ENV	SOC	GOV	ESG	ENV	SOC	GOV
2007	-0.1634	-0.1832	-0.0782	-0.132	-0.0455	-0.0502	-0.0479	0.0145	0.0036	0.0007	0.0042	-0.0787
2008	-0.0477	0.0081	-0.0961	-0.0888	-0.0532	-0.057	-0.0204	0.0249	-0.0118	-0.0451	0.0164	0.0582
2009	-0.0426	-0.0508	-0.017	-0.0673	-0.0267	-0.0258	-0.0264	-0.0376	-0.0328	-0.0562	-0.0731	-0.013
2010	0.0364	0.0726	0.0093	-0.0856	-0.1013*	-0.1089*	-0.0469	-0.0641	0.0237	0.0275	0.0306	0.0123
2011	0.104	0.1334*	0.079	0.0832	-0.0461	-0.0426	-0.0431	-0.0399	-0.045	-0.0083	-0.0603	-0.0532
2012	0.029	0.0797	0.0008	0.0095	-0.0148	-0.0195	0.0302	-0.0621	-0.0822	-0.099	-0.0906	-0.0198
2013	0.0724	0.0948	0.0823	0.0443	-0.0943*	-0.0771*	-0.1131*	-0.0595	-0.0851	-0.0806	-0.088	-0.1179*
2014	-0.0528	-0.0671	-0.0157	-0.0277	-0.1346*	-0.1254*	-0.1045*	-0.1023*	-0.1079*	-0.1113*	-0.0548	-0.1505*
2015	0.0297	0.0263	0.031	-0.0227	-0.0391	-0.0272	-0.0224	-0.059	-0.0256	-0.0337	-0.0294	-0.0277
2016	-0.0975	-0.1087*	-0.091	-0.058	-0.042	-0.0424	-0.0292	-0.0526	-0.0714	-0.0795*	-0.0687	-0.0109
2017	-0.0692	-0.0956	-0.0048	-0.0555	-0.1496*	-0.1449*	-0.1201*	-0.1455*	-0.1048*	-0.1083*	-0.0672	-0.1026*

Panel C

-	ESG	ENV	SOC	GOV	ESG	ENV	SOC	GOV	ESG	ENV	SOC	GOV
2007	0.0475	0.0856	0.0305	-0.1441	-0.0707	-0.0950*	-0.0935	0.0125	0.189	0.1889	0.096	0.191
2008	-0.2003*	-0.1681*	-0.2382*	-0.1299	-0.0745*	-0.0603	-0.0391	-0.0720*	0.0104	-0.0192	0.0222	0.0969
2009	-0.0507	-0.064	-0.0139	-0.0345	-0.1349*	-0.1384*	-0.0895*	-0.0521	0.0188	0.0073	-0.0201	0.0395
2010	0.0182	-0.0151	0.012	0.0877	-0.0261	-0.0144	-0.0135	-0.0703*	0.0072	-0.0161	-0.0627	0.0265
2011	0.0694	0.1285*	-0.0113	0.0252	-0.0993*	-0.0709*	-0.1129*	-0.1357*	-0.0641	-0.0641	-0.0351	-0.071
2012	-0.0431	-0.0258	-0.0195	-0.0636	-0.0618	-0.0509	-0.0747*	-0.0484	-0.0131	0.0041	-0.0247	-0.0295
2013	-0.0024	0.0255	0.0012	-0.0267	-0.1266*	-0.1068*	-0.1222*	-0.1001*	-0.0589	-0.035	-0.0899	-0.0402
2014	-0.0467	-0.0442	0.0024	-0.073	-0.0932*	-0.0876*	-0.0602	-0.0834*	-0.0025	-0.0069	-0.0139	-0.0331
2015	0.0297	0.0263	0.031	-0.0227	-0.0678*	-0.0648*	-0.0356	-0.0441	-0.0534	-0.0633	-0.0466	-0.0840*
2016	0.1458*	0.1270*	0.1041*	0.1259*	0.0462	0.0486	0.0452	0.0128	-0.0493	-0.0447	-0.0634	-0.0373
2017	-0.035	-0.0481	0.0154	0.025	-0.0423	-0.0448	-0.0257	-0.0590*	-0.0635	-0.0596	-0.0596	-0.0533

Overall, the risk analysis employed in this paper does not lend much support to the idea that there is a relationship between the level of responsibility and risk moments. Volatility negatively correlates with ESG scores in the US and to a lesser degree in Europe. Results for higher moments of risk (skewness and kurtosis) are very sporadically significant.

4.3 Robustness Test

Appendix 4 reports the regression results of the difference portfolios, based on a segmentation of the whole sample in the 20^{th} (irresponsible) and 80^{th} (responsible) percentile. The regression outputs largely confirms the findings of the 30^{th} and 70^{th} percentile portfolios. All alphas, no matter the model specification, remain insignificant in Europe. In Japan, the CAPM alphas disappear but the insignificance of alphas in the other model specifications also remains. In the US, the significant alphas in both the *ESG* and *GOV* dimension remain and even slightly grow (yearly excess return): Investing in the US portfolio that contains the assets with the lowest 20% of all *ESG* composite scores (*GOV* scores) in each year and shorting the portfolio with the highest 20% of *ESG* composite scores (*GOV* scores) would have yielded an excess return of 0.78% (10.41%) per month, which corresponds to a yearly overperformance of 9.8% (10%) in these two investment segments. Explanatory power for these two segments remains low.

More insightful are the regressions results of the equal-weighted portfolios. Since this robustness test especially aims at identifying a potential bias in the sample which leads to the consistently positive alphas in the long-only portfolios, coefficients of the the difference portfolios and the long-only portfolios are reported in appendix 5.

Two observations stand out: First of all, allegedly positive alphas in the long-only portfolios identified earlier in this study lose their significance. Looking at the 6-factor model, alphas in almost half of the investment segments become insignificant. Whenever alphas remain significant, they decrease significantly across all model specifications. These findings support the idea of potential overperformers in the ESG sample which are overaccentuated in the value-weighted portfolios, whereas they lose some of their ability to influence the overall portfolio performance in the equal-weighted portfolio. Secondly, R^2 slightly decreases in many cases. This is not surprising due to the fact that the common risk factors from Kenneth French's data library are built in a value-weighted process, thus they are a better fit to explain value-weighted portfolio returns.

4.4 COMBINED EMPIRICAL RESULTS

The analysis shows that the notion of compensation effects is misleading in most investment segments, as higher ESG scores hardly have to compensate for lower returns and/or higher risk.

Time-series monthly excess return averages are mostly higher for irresponsible portfolios, though the means do not always differ significantly from zero. Volatility tends to be lower for responsible portfolios and is also confirmed on stock level, especially in the US and to a lesser extent in Europe. Nonetheless, on portfolio level the effect of the high mean excess returns outweighs the volatility effect as reflected in the Sharpe Ratios which are almost always higher in the irresponsible portfolios. When higher moments of risk are taken into account, the Adjusted Sharpe Ratio still makes a strong statement in favor of irresponsible portfolios, especially in the investment segments Japan *ESG* and US *ESG* and *GOV* where the ratio is above 1.

Testing the difference between the excess returns for responsible and irresponsible portfolios, only the difference portfolios in the *ESG* and *GOV* dimensions in the US are able to produce a significant alpha after controlling for 6 risk factors that are commonly used to explain stock returns: While each pair of long-only portfolios yields a significantly positive alpha, the outperformance of the respective irresponsible portfolios is so large only in the US *ESG* and *GOV* dimension that shorting the portfolios containing the 30% most responsible assets and buying the portfolio with the 30% least responsible assets yields a significant excess return of annually 8% (*ESG*) and 10% (*GOV*) (with the breakpoints at the 20^{th} and 80^{th} percentile, the yearly excess return is 9.8% for *ESG* and 10% for *GOV*). T-statistics about the average mean excess returns also suggest an overperformance of the irresponsible portfolio in Japan's *ENV* dimension and the CAPM supports this idea, however, these intercepts disappear when adding the size, value and momentum factor, as well as the profitability and investment factor. The alleged overperformance that the CAPM detects is thus explained by a linear combination of these risk factors and has nothing to do with the nature of the sort variable, that is, ESG scores.

Referring to the hypotheses presented in chapter 2.4, the lack of significance of the alphas of most difference portfolios that go long (short) in stocks with the best (worst) ESG ratings does not allow to reject the hypothesis (1) that there is no difference in the returns, adjusted for common risk factors, between socially responsible and irresponsible stocks as measured by their ESG scores. With reference to hypothesis 3 and 4, the only exceptions to these findings are the *ESG* and *GOV* dimensions in the US. The correlation analysis does not permit to reject the hypothesis (2), that is no significant correlation between socially responsible stocks and higher risk as expressed by negative skewness and excess kurtosis.

On portfolio level, the hypothesis (2) can be rejected as higher risk tends to be lower in Europe's responsible portfolios in the *ESG*, *SOC* and *GOV* dimensions and kurtosis tends to be lower in the US' responsible portfolios in the *ESG*, *ENV* and *SOC* dimensions. Whereas the study largely fails to detect a difference between responsible and irresponsible assets, the results (positive alphas in all long-only portfolios) suggest that assets that are rated by Bloomberg might be financially strong stocks in the first place, which could explain the outperformance in relation to the overall market. These results are accentuated in the value-weighted portfolios, suggesting the existence of some very large (by market capitalization) high-performing stocks in the sample. The results are therefore less accentuated in equal-weighted portfolios.

With regards to potential compensation effects, the study reveals that in all investment segments, except US *ESG* and *GOV*, responsible investors can equally enjoy the feeling of holding assets with high ESG scores and benefit from comparably (high!) returns. Put differently, only in the US *ESG* and *GOV* segments, both the significant alpha in the long-short strategy and the Adjusted Sharpe Ratio support the idea that investors in the responsible portfolios lose utility in a suboptimal return-risk profile compared to the irresponsible portfolio (even if they still outperform the market). The results from this study suggest that in all other cases, responsible investors do not suffer from lower returns or higher risk in their portfolios compared to irresponsible portfolios. They do not have to be "compensated" by the non-financial utility gained from the responsible nature of their assets, but this non-financially utility is rather "the cherry on top" of a financial outperformance. The next chapter is dedicated to suggest some possible explanations for these findings, mostly based on the methodology employed in this paper.

5. Discussion of Results and Limitations

The wide lack of significance in alphas of most difference portfolios and the generally weak association between ESG scores and moments of risk, leading to a general rejection of the idea of "compensation effects" in SRI urges to make some statements related to the methodology and the results of this study.

The results from this study largely confirm the idea that there is no relation between the level of responsibility of an asset and its financial return, as presented with the idea of efficient markets in chapter 2.2. Also, it should be noted that Halbrittner and Dorfleitner (2015) have troubles identifying any relation between Bloomberg's ESG scores and financial returns both on portfolio- and stock level. As stated earlier, the findings of SRI studies seem to depend very much on the chosen methodology. Approximating the idea of SRI assets by using Bloomberg scores and applying multifactor models to assess their financial performance turns out to bear the following caveats:

The first problem lies in the quality of Bloomberg's ESG scores. Numerous assets are rated with the same scores throughout years, especially in the *GOV* dimension. While this problem was methodologically alleviated by building portfolios instead of doing full-sample panel data regression, the question remains whether the ratings are reliable and actually updated yearly based on the latest information reflecting the actual level of sustainability of a given asset.

Moreover, the methodology of this study centrally rests on the assumption that Bloomberg publishes ESG ratings in the most comprehensive way, i.e. for as many companies as possible and independently of firm characteristics that might otherwise be related to the financial performance of the firm. In other words, there should be no relationship between the financial performance of a company and the likelihood that Bloomberg publishes the company's scores. However, it is possible that Bloomberg tends to rate those firms that are financially successful historically (as indicated by the consistently positive alphas in the long-only portfolios). It is important to keep in mind that the assets used in this study represent a sub-sample of the market, but the overall market is used to build the risk factors to evaluate the portfolio performance, notably the market risk factor *MKT*. While this is common practice in similar performance studies, there is a risk of yielding misleading results if the sub-sample tends to contain financial "overachievers" for structural reasons. If extremely positive outliers in comparison to the overall market are found in the portfolios in this study, they can easily increase the overall portfolio performance, all the more if they have a large market capitalization in the value-weighted approach. The robustness check using equal-weighted portfolios supports the idea, as extremely positive outliers with a large market capitalization do not receive such extreme weights in the portfolio, leading to

generally smaller alphas. To further improve the methodology of the paper and to check if the suspicion is correct, the risk factors for the performance analysis should be built based on the actual sample instead of relying on the imperfect risk proxies from Kenneth French's data library.

Equally importantly, the meaning and information content of Bloomberg's ESG scores must be questioned. The investors in this study consider responsibility only on the basis of an asset's ESG score by Bloomberg, which is essentially a transparency score. Companies in inherently "critical" industries might still get a good rating if they operate comparatively responsible in their sector and are very willing and active in sharing their sustainability efforts. Since in certain industries their overall perception of responsibility might be bad (e.g. mining industry and environment), the logic of the scores seems somewhat counter-intuitive and is probably not what many SRI investors have in mind when picking their stocks based on sustainability scores. This aspects about the true motives leads to another critical point. As has been pointed out in section 2.2.1, SRI investors' motives can vary from truly intrinsic motivation for social change, to rather reputation-based (signaling) aspects and regulatory restrictions. If investing in a socially responsible manner is mostly done for signaling reasons or due to legal restriction without a real interest to change the world for the better (i.e. allegedly SRI investors gain non-financial utility from signaling or from fulfilling certain requirements), then the information content of Bloomberg's ESG scores as a mere indicator of transparency is not as critical. The most important aspect for these types of SRI investors is that ESG scores prove their adherence to social or regulatory investment standards. The same is true for investors who derive utility from the transparency or information disclosure that ESG scores promise, as emphasized by Puaschunder (2017). For SRI investors who have the true intrinsic motivation to achieve positive social change, in one or the other dimension of sustainability, the scores provided by Bloomberg might be useless. Especially since this type of SRI investors tends to be motivated by the "perceived effectiveness" of their investment, the nature of Bloomberg's ESG rating is not effective to incorporate these scores in the investment process. Keeping this in mind, the identification of "compensation effects", in at least two investment segments in this study, becomes questionable when doubting that intrinsically motivated SRI investors can actually gain non-financial utility from Bloomberg scores.

6. Conclusion

"SRI has become a multidimensional concept that now serves the needs of a heterogeneous group of investors who differ in the pecuniary and non-pecuniary benefits that they expect to derive from their investment" (Derwall et al., 2011, p. 2145). Other than in the classic finance theory, the idea of SRI investments is based on the notion that investors derive their utility from other factors than only risk and return. Altruistic, genuine interest in positive social change, pro-social behavior (merely signaling a dedication towards sustainability issues) or regulatory restrictions equally drive investment decisions and are expected to become even more important in the future, as the awareness of the financial world's responsibility towards sustainability issues increases. To what extent these non-financial utility aspects have to compensate for an allegedly suboptimal risk-return profile is the subject of this study.

Considering three geographies (Europe, Japan and the US) and four distinct measures of sustainability (Bloomberg's ESG composite score, as well as the environmental, governance and social subscores), the resulting 12 investment segments were scrutinized to detect patterns in returns and risk and to make a statement if the sustainable nature of more responsible assets, as measured by the highest ESG scores, has to compensate for a suboptimal risk/return profile. The results widely reject the idea of compensation effects, due to a lack of significance in alphas of most difference portfolios (which allow to establish a link between the sorting variable, ESG scores, and the outcome variable, excess returns) and a weak association between ESG scores and moments of risk. Only in two investment segments, that is in the US *ESG* and *GOV* dimensions, shorting the responsible portfolios and going long in the irresponsible portfolios yields a significant excess return that can not be explained by common risk factors or higher moments of risk. However, looking at long-only portfolios, the financial performance is superior to the overall market in practically all investment segments, depending on the methodology (i.e. the cut-offs for portfolio construction and asset weights). It seems reasonable to assume that the selection of stocks that are rated with ESG scores by Bloomberg is not entirely random, and that there might be an undetected link between the existence of a ESG score and a financial overperformance in comparison with the overall market. Therefore, the results in this study should be considered together with some suggestions for methodological improvements, notably related to the overall market as a benchmark to evaluate the performance of assets with high versus low Bloomberg ESG scores.

Finally, the results should be considered together with the true meaning of Bloomberg's ESG scores, i.e. the ability or disability of these scores to express the true sustainable nature of the stock. Bloomberg's ESG scores being mere transparency scores, not all types of SRI investors, notably not the type of investor who seeks to achieve genuine social change,

might be able to deduct non-financial utility from investing in assets with high Bloomberg ESG scores. For SRI investors who are prone to show pro-social behavior or who are restricted by regulation, however, mixing money and “morality” by investing in stocks with high Bloomberg ESG scores appears to be a rewarding strategy.

LITERATURE

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

Bali, T. G., Engle, R. F., & Murray, S. (2016). *Empirical Asset Pricing: The Cross-Section of Stock Returns*. Hoboken, NJ: Wiley.

Bauer, R., Koedijk, K., & Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. *Journal of Banking and Finance*, *29*(7), 1751–1767.

Bauer, R., & Smeets, P. (2015). Social identification and investment decisions. *Journal of Economic Behavior and Organization*, *117*, 121–134.

Beal, D. J., Goyen, M., & Philips, P. (2005). Why Do We Invest Ethically? *The Journal of Investing*, *14*(3), 66–78.

Becker, G. (1957). *The Economics of Discrimination*. Chicago: University Press.

Belghitar, Y., Clark, E., & Deshmukh, N. (2014). Does it pay to be ethical? Evidence from the FTSE4Good. *Journal of Banking and Finance*, *47*(1), 54–62. Retrieved from http://dx.doi.org/10.1016/j.jbankfin.2014.06.027

Blitz, D., & Fabozzi, F. J. (2017). Sin Stocks Revisited: Resolving the Sin Stock Anomaly. *The Journal of Portfolio Management*, *44*(1), 1–7.

Bloomberg Finance L.P. (2014). Look Beyond. Retrieved from https://cfaboston.org/docs/ESG/BloombergLookBeyond2014.pdf

Bloomberg Finance L.P. (2019). Sustainable Finance: Bloomberg Professional Services. Retrieved from https://www.bloomberg.com/professional/solution/sustainable-finance/

Bodie, Z., Kane, A., & Marcus, A. J. (2014). *Investments* (10th ed.). New York: McGraw-Hill Education.

Bollen, N. P. B. (2007). Mutual Fund Attributes and Investor Behavior. *The Journal of Financial and Quantitative Analysis*, *42*(3), 683–708.

Borgers, A., Derwall, J., Koedijk, K., & ter Horst, J. (2015). Do social factors influence investment behavior and performance? Evidence from mutual fund holdings. *Journal of Banking and Finance*, *60*, 112–126.

Brodback, D., Guenster, N., & Mezger, D. (2018). Altruism versus Egoism in Investment Decisions. *Review of Financial Economics*, *37*(1), 118–148.

Cojuharenco, I., Cornelissen, G., & Karelaia, N. (2016). Yes, I can: Feeling Connected to Others Increases Perceived Effectiveness and Socially Responsible Behavior. *Journal of Environmental Psychology*, *48*, 75–86.

Columbia University Libraries. (2009). Company Information Guide: ESG Info.

Connaker, A., & Madsbjerg, S. (2019). The State of Socially Responsible Investing. *Harvard Business Review*.

De, I., & Clayman, M. R. (2015). The Benefits of Socially Investing: An Active Manager's Perspective. *The Journal of Investing*, 49–72.

Derwall, J. (2007). *The Economic Virtues of SRI and CSR*. Erasmus University Rotterdam: Erasmus Research Institute of Management.

Derwall, J., Koedijk, K., & ter Horst, J. (2011). A tale of values-driven and profit-seeking social investors. *Journal of Banking and Finance*, *35*(8), 2137–2147.

Drempetic, S., Klein, C., & Zwergel, B. (2019). The Influence of Firm Size on the ESG Score:

Corporate Sustainability Ratings Under Review. *Journal of Business Ethics*, 1–28.
Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Retur. *The Journal of Finance*, *47*(2), 427–465.
Fama, E. F., & French, K. R. (2007). Disagreement, Tastes, and Asset Prices. *Journal of Financial Economics*, *83*(3), 667–689.
French, K. R. (2019). Current Research Returns. Retrieved from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance and Investment*, *5*(4), 210–233.
Friedman, M. (1970). The Social Responsibility of Business Is to Increase Its Profits. *The New York Times Magazine*.
Gibson, R., Krueger, P., Matos, P., & Steffen, T. (2019). Responsible Institutional Investing Around the World.
GSIA. (2019). *Global Sustainable Investment Review 2018*. Retrieved from http://www.gsi-alliance.org/trends-report-2018/
Halbritter, G., & Dorfleitner, G. (2015). The wages of social responsibility - where are they? A critical review of ESG investing. *Review of Financial Economics*, *26*, 25–35.
Heinkel, R., Kraus, A., & Zechner, J. (2001). The Effect of Green Investment on Corporate Behavior. *The Journal of Financial and Quantitative Analysis*, *36*(4), 431–449.
Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, *93*(1), 15–36.
Huppé, G. A. (2011). Alpha’s Tale: The Economic Value of CSR. In *PRI Academic Network Conference*.
Investopedia. (2019). Efficient Market Hypothesis (EMH) Definition. Retrieved from https://www.investopedia.com/terms/e/efficientmarkethypothesis.asp
Kempf, A., & Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management*, *13*(5), 908–922.
Manescu, C. (2011). Stock returns in relation to environmental, social and governance performance: Mispricing or compensation for risk? *Sustainable Development*, *19*(2), 95–118.
Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91.
Oikonomou, I., Brammer, S., Brooks, C., & Pavelin, S. (2012). Corporate Social Performance and Stock Returns: UK Evidence from Disaggregate Measures. *Journal of Financial Management*, *4*(12), 483–515.
Pezier, J., & White, A. (2006). The Relative Merits of Investable Hedge Fund Indices and of Funds of Hedge Funds in Optimal Passive Portfolios. In *ICMA Centre Discussion Papers in Finance*. Henley Business School, Reading University.
PRI. (2018). *Annual Report 2018*. Retrieved from https://www.unpri.org/Uploads/g/f/c/priannualreport_605237.pdf
PRI. (2019). What is Responsible Investment? Retrieved from https://www.unpri.org/pri/what-is-responsible-investment
Puaschunder, J. M. (2017). Socio-Psychological Motives of Socially Responsible Investors. *Advances in Financial Economics*, *19*, 209–247.
Renneboog, L., ter Horst, J., & Zhang, C. (2007). *The Price of Ethics: Evidence from Socially Responsible Mutual Funds*. Tilburg.
Riedl, A., & Smeets, P. (2017). Why Do Investors Hold Socially Responsible Mutual Funds?

Journal of Finance, *72*(6), 2505–2550.

Salaber, J. (2007). *The Determinants of Sin Stock Returns. Evidence on the European Market*. University of Paris Daufin.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, *19*(3), 425–442.

Statman, M. (2005). Normal Investors, Then and Now. *Financial Analysts Journal*, *61*(2), 31–37.

Statman, M. (2008). *The expressive nature of socially responsible investors*. Santa Clara University, Leavey School of Business.

Utz, S., & Wimmer, M. (2014). Are they any good at all? A financial and ethical analysis of socially responsible mutual funds. *Journal of Asset Management*, *15*(1), 72–82.

APPENDIX

APPENDIX 1: DETAILED ESG SUMMARY STATISTICS

The table reports the number of observations (Obs), the mean, the standard deviation (SD) as well as the minimum (Min) and maximum (Max) value of the ESG sample per year, for each investment segment.

			2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
EU	**ESG**	**Obs**	123	220	251	266	312	350	365	382	392	416	422
		Mean	28.00	28.23	30.08	32.44	32.10	31.97	33.56	35.53	37.35	37.85	39.20
		SD	13.22	13.59	13.88	14.80	16.50	17.52	17.36	16.27	15.43	15.10	14.76
		Min	0.83	1.24	2.89	2.89	3.31	4.13	3.31	3.31	3.31	3.31	5.37
		Max	54.55	58.26	61.98	64.46	73.14	76.86	80.58	78.51	70.25	67.36	68.60
	ENV	**Obs**	104	198	227	246	258	276	294	331	361	385	396
		Mean	25.02	23.44	25.72	28.00	30.08	31.18	32.60	32.52	32.47	32.69	33.98
		SD	13.97	15.14	15.37	16.61	16.70	17.21	16.76	16.39	16.49	15.92	15.76
		Min	2.33	2.33	2.33	1.55	1.55	1.55	2.33	2.33	2.33	1.55	2.33
		Max	52.71	58.91	64.34	73.64	75.97	80.62	84.50	82.95	70.54	68.99	69.77
	SOC	**Obs**	120	218	249	262	278	298	316	347	372	403	410
		Mean	29.43	29.31	31.10	33.20	35.74	37.85	39.71	40.93	42.15	42.46	43.16
		SD	16.83	17.56	18.27	18.34	18.07	18.52	17.21	16.60	16.55	16.94	16.69
		Min	3.13	3.13	3.13	3.33	3.33	3.33	3.33	3.33	3.51	3.51	3.51
		Max	61.67	66.67	75.44	77.19	80.70	80.70	75.44	82.46	78.95	78.95	78.95
	GOV	**Obs**	120	219	251	266	312	350	365	382	392	416	422
		Mean	43.53	43.61	44.80	47.38	49.08	49.20	49.99	51.03	52.08	52.33	53.50
		SD	12.90	13.13	13.10	13.36	14.24	14.21	13.78	13.47	12.70	12.51	11.81
		Min	3.57	8.93	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29
		Max	71.43	75.00	73.21	73.21	85.71	85.71	85.71	83.93	85.71	80.36	76.79
Japan	**ESG**	**Obs**	634	1094	1476	1508	1650	1725	1744	1777	1785	1784	1769
		Mean	21.24	22.31	19.33	20.90	20.87	20.75	21.57	21.74	22.37	22.74	23.88
		SD	12.81	12.30	12.48	12.83	12.85	12.75	12.58	12.72	12.60	12.60	12.35
		Min	2.89	4.13	3.31	2.48	0.96	2.48	2.89	2.89	3.31	3.31	4.13
		Max	50.83	53.72	52.48	58.26	57.02	58.68	60.33	60.74	60.74	60.74	62.81
	ENV	**Obs**	428	781	841	876	908	933	1026	1042	1099	1118	1237
		Mean	25.06	26.30	26.09	26.44	26.54	26.20	25.04	25.12	24.46	24.30	23.22
		SD	16.81	15.25	15.30	15.50	15.72	15.68	15.97	16.07	16.34	16.54	16.85
		Min	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78
		Max	56.59	58.91	58.91	67.44	65.12	68.22	72.87	74.42	74.42	74.42	78.29
	SOC	**Obs**	581	935	1010	1032	1061	1082	1176	1193	1360	1507	1657
		Mean	17.73	18.13	18.25	18.88	19.27	19.60	19.70	20.49	19.76	19.08	19.12
		SD	11.10	11.17	11.27	11.49	11.63	11.56	11.64	11.83	11.84	12.12	11.84
		Min	3.13	3.13	3.13	3.13	3.13	3.13	3.13	3.13	3.13	3.13	3.33
		Max	57.81	54.39	54.39	59.65	59.65	64.91	64.91	64.91	59.65	64.91	63.16
	GOV	**Obs**	634	1094	1475	1508	1650	1725	1744	1777	1785	1784	1769
		Mean	36.27	37.10	36.40	41.62	43.74	44.26	45.47	45.72	46.35	46.46	47.26
		SD	8.48	7.61	7.71	7.64	6.54	5.95	4.97	4.89	4.63	4.68	4.50
		Min	8.93	12.50	12.50	10.71	3.57	10.71	12.50	12.50	12.50	12.50	12.50
		Max	57.14	57.14	57.14	62.50	62.50	67.86	67.86	67.86	67.86	67.86	67.86
US	**ESG**	**Obs**	401	1039	1308	1631	2552	2735	2852	3028	3172	3324	3376
		Mean	19.09	16.83	16.97	16.94	15.56	15.84	16.31	16.60	17.02	17.63	18.24
		SD	10.14	9.11	9.65	10.01	9.11	9.57	9.91	10.21	10.31	10.32	10.41
		Min	1.24	0.83	0.83	0.83	1.32	1.32	3.07	3.51	2.19	1.65	2.19
		Max	57.44	61.16	64.46	72.61	73.44	75.10	72.73	74.79	75.62	70.25	72.31

ENV	**Obs**	169	339	442	544	610	674	741	809	924	1181	1357
	Mean	17.75	16.04	16.15	16.62	17.40	18.22	18.68	19.15	18.29	15.81	14.90
	SD	13.63	14.44	15.47	16.34	16.49	16.95	17.16	17.35	17.41	16.91	16.59
	Min	0.78	1.38	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78
	Max	51.94	55.37	66.67	77.69	77.69	78.51	76.74	82.17	75.97	78.29	74.42
SOC	**Obs**	361	808	1021	1151	1339	1421	1591	1724	1899	2305	2493
	Mean	13.63	11.76	12.16	13.02	13.36	14.43	15.42	16.25	16.93	16.67	17.09
	SD	13.18	12.06	12.35	12.82	12.70	13.10	13.16	13.15	13.07	12.64	12.22
	Min	3.13	3.13	3.13	3.13	3.13	3.13	3.13	3.13	3.13	3.13	3.13
	Max	73.44	68.75	73.44	76.56	78.33	78.33	86.67	73.68	79.69	85.96	77.19
GOV	**Obs**	400	1035	1304	1630	2551	2735	2852	3027	3172	3323	3375
	Mean	51.95	50.70	50.54	50.54	49.93	49.93	49.91	49.96	50.42	50.87	51.48
	SD	6.68	6.65	7.17	7.15	6.50	6.63	6.82	6.98	6.68	6.74	6.66
	Min	5.36	3.57	3.57	3.57	3.57	3.57	3.57	3.57	3.57	3.57	3.57
	Max	71.43	73.21	76.79	78.57	83.93	82.14	85.71	82.14	85.71	85.71	85.71

APPENDIX 2: ANNUALIZED RETURNS

The table present the annualized excess returns of all 36 portfolios (irresponsible, responsible, difference) per year, for each investment segment.

			2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
EU	**ESG**	**irresp.**	12.23%	-37.99%	27.87%	57.74%	-20.19%	41.91%	31.45%	1.96%	9.27%	25.34%	8.84%
		resp.	7.11%	-22.76%	4.44%	44.68%	-22.51%	45.63%	27.19%	-3.05%	-2.91%	26.58%	7.95%
		diff.	1.36%	-17.07%	21.93%	8.42%	1.58%	-2.87%	2.21%	4.99%	11.43%	-1.07%	-0.40%
	ENV	**irresp.**	-1.13%	-37.40%	22.20%	55.59%	-26.73%	47.25%	27.09%	1.55%	4.92%	28.31%	6.12%
		resp.	8.54%	-21.78%	3.40%	45.51%	-22.02%	46.92%	26.30%	-4.15%	-2.78%	25.39%	8.96%
		diff.	-12.19%	-17.82%	18.83%	6.74%	-6.10%	0.11%	-0.25%	5.85%	6.69%	2.25%	-3.84%
	SOC	**irresp.**	-1.61%	-32.89%	23.93%	56.81%	-20.81%	41.25%	25.05%	2.36%	7.94%	27.93%	14.20%
		resp.	3.34%	-22.19%	2.89%	45.36%	-23.88%	46.16%	26.20%	-2.94%	-3.31%	26.42%	8.53%
		diff.	-8.21%	-12.59%	19.26%	6.70%	2.19%	-3.86%	-1.71%	5.27%	10.54%	0.96%	3.90%
	GOV	**irresp.**	-5.92%	-37.02%	22.43%	54.63%	-24.21%	31.13%	28.25%	-1.35%	7.70%	24.82%	10.99%
		resp.	-0.04%	-31.08%	6.60%	37.65%	-24.18%	48.79%	27.55%	-1.95%	-1.25%	24.74%	5.01%
		diff.	-9.08%	-6.83%	15.14%	11.21%	0.39%	-12.60%	-0.49%	0.47%	8.23%	-0.30%	4.45%
Japan	**ESG**	**irresp.**	0.39%	-8.26%	15.93%	23.51%	8.68%	35.82%	13.16%	17.34%	18.74%	19.99%	22.39%
		resp.	-2.10%	-17.61%	10.43%	14.24%	-5.92%	33.37%	10.83%	22.95%	-0.66%	20.06%	12.38%
		diff.	-1.54%	9.23%	4.49%	7.70%	13.89%	2.26%	2.21%	-4.79%	18.16%	-0.56%	7.56%
	ENV	**irresp.**	-3.13%	-12.37%	20.48%	21.36%	8.55%	40.95%	17.63%	20.06%	5.29%	24.64%	21.33%
		resp.	-2.10%	-17.61%	10.43%	14.24%	-5.92%	33.37%	10.83%	22.95%	-0.66%	20.06%	12.38%
		diff.	-1.54%	9.23%	4.49%	7.70%	13.89%	2.26%	2.21%	-4.79%	18.16%	-0.56%	7.56%
	SOC	**irresp.**	-3.08%	-15.52%	14.78%	21.49%	6.97%	33.06%	18.25%	24.51%	4.71%	24.98%	14.47%
		resp.	0.96%	-18.32%	8.82%	14.44%	-8.25%	36.35%	10.90%	21.94%	-0.25%	19.59%	12.17%
		diff.	-6.93%	1.70%	5.27%	6.08%	15.89%	-2.44%	6.78%	2.19%	4.58%	4.09%	0.84%
	GOV	**irresp.**	1.50%	-16.40%	23.64%	23.78%	3.55%	28.64%	10.43%	12.31%	10.10%	18.76%	25.44%
		resp.	-4.20%	-19.01%	9.87%	12.96%	-6.92%	35.44%	10.75%	20.99%	-1.64%	19.12%	10.57%
		diff.	2.69%	2.11%	12.60%	9.05%	10.59%	-5.08%	-0.28%	-7.44%	10.77%	-0.92%	12.04%
US	**ESG**	**irresp.**	11.72%	-22.46%	46.93%	50.65%	5.54%	49.11%	39.20%	24.35%	-4.09%	36.38%	25.52%
		resp.	0.70%	-22.99%	28.19%	28.22%	3.79%	30.60%	24.74%	15.22%	5.17%	22.14%	14.34%
		diff.	6.87%	0.37%	14.39%	18.13%	1.69%	14.64%	11.89%	8.10%	-6.46%	11.66%	8.75%
	ENV	**irresp.**	-8.00%	-32.46%	48.30%	35.71%	4.07%	35.08%	24.26%	17.08%	17.04%	19.31%	13.55%
		resp.	-2.36%	-22.43%	26.25%	24.71%	5.34%	27.37%	21.20%	14.23%	2.52%	22.10%	13.17%
		diff.	-8.77%	-12.56%	17.56%	8.37%	-1.14%	6.15%	2.49%	2.36%	14.30%	-2.61%	-0.72%
	SOC	**irresp.**	3.50%	-32.97%	59.53%	18.30%	5.29%	43.85%	34.19%	15.83%	12.70%	28.00%	20.21%
		resp.	-0.43%	-23.12%	23.28%	28.95%	-0.90%	32.73%	20.58%	12.02%	2.02%	21.99%	14.36%
		diff.	0.44%	-10.03%	31.08%	-8.23%	6.34%	8.51%	11.46%	3.30%	10.18%	4.81%	3.95%
	GOV	**irresp.**	13.57%	-24.42%	42.49%	30.23%	3.59%	46.80%	43.50%	21.08%	-3.28%	45.81%	33.41%
		resp.	-1.88%	-23.04%	29.90%	25.92%	4.65%	29.19%	22.37%	14.50%	2.48%	20.88%	12.28%
		diff.	12.49%	-2.66%	9.19%	1.91%	0.21%	14.05%	17.86%	5.60%	-4.02%	20.36%	18.16%

APPENDIX 3: AVERAGE MONTHLY EXCESS RETURNS AND T-TESTS

The table reports the time-series averages for the 132 monthly excess returns of all 36 portfolios. T-tests are performed and reported to test the hypothesis that the mean excess returns differ statistically from zero, indicating a cross-sectional relation between ESG scores and excess returns. The asterisk (*) indicates statistical significance at the 95% level (t-statistics > 2.00). Minimum (Min) and maximum (Max) monthly excess returns are also reported.

			average monthly excess return	t-test	Min	Max
EU	ESG	irresp.	1.10%	1.9336	-26.17%	22.35%
		resp.	0.83%	1.5832	-12.88%	14.99%
		diff.	0.22%	0.8972	-13.95%	10.29%
	ENV	irresp.	0.88%	1.5043	-27.20%	20.86%
		resp.	0.84%	1.6136	-13.62%	15.05%
		diff.	-0.01%	-0.0290	-16.66%	11.19%
	SOC	irresp.	1.01%	1.8952	-22.87%	18.83%
		resp.	0.80%	1.4710	-13.26%	16.28%
		diff.	0.17%	0.7626	-10.47%	6.02%
	GOV	irresp.	0.79%	1.3780	-25.41%	20.53%
		resp.	0.68%	1.2351	-22.05%	15.53%
		diff.	0.07%	0.3522	-4.61%	6.31%
Japan	ESG	irresp.	1.25%*	3.4591	-13.30%	15.09%
		resp.	0.76%*	2.1267	-10.93%	11.48%
		diff.	0.44%	1.7210	-9.18%	10.65%
	ENV	irresp.	1.23%*	3.2148	-11.19%	12.67%
		resp.	0.79%*	2.2776	-11.06%	10.75%
		diff.	0.39%*	2.1057	-5.21%	7.56%
	SOC	irresp.	1.09%*	2.8115	-10.27%	13.21%
		resp.	0.76%*	2.0875	-12.24%	12.32%
		diff.	0.28%	1.4489	-6.75%	5.45%
	GOV	irresp.	1.08%*	2.8595	-11.56%	13.65%
		resp.	0.69%	1.8913	-11.01%	12.25%
		diff.	0.34%	1.6432	-7.17%	7.29%
US	ESG	irresp.	1.81%*	3.5417	-16.45%	26.98%
		resp.	1.11%*	2.8747	-16.10%	12.02%
		diff.	0.66%*	2.5426	-8.46%	17.57%
	ENV	irresp.	1.2%*	2.6865	-14.14%	18.72%
		resp.	0.98%*	2.6540	-15.16%	11.47%
		diff.	0.17%	0.7883	-6.88%	14.13%
	SOC	irresp.	1.5%*	2.5931	-21.05%	39.26%
		resp.	0.97%*	2.6024	-16.09%	11.73%
		diff.	0.48%	1.3359	-16.15%	35.18%
	GOV	irresp.	1.77%*	3.4854	-13.47%	20.08%
		resp.	1.01%*	2.6842	-15.49%	11.30%
		diff.	0.71%*	2.4252	-6.77%	11.32%

APPENDIX 4: REGRESSION RESULTS FOR THE DIFFERENCE PORTFOLIOS (20TH/80TH PERCENTILES)

The table presents the risk-adjusted alphas (intercept), factor sensitivities for the risk factors size (SMB), value (HML), momentum (MOM), profitability (RMW) and investment (CMA), and the adjusted R^2 for the 12 zero-cost difference portfolios which are long in the irresponsible and short in the responsible portfolios based on the $20^{th}/80^{th}$ percentiles. For each investment segment, outputs of the three model specifications (I)-(III) are reported. The asterisk (*) indicates statistical significance at the 95% level.

		EU			Japan			US		
		(I) CAPM	**(II) Carhart**	**(III) 6 Factor**	**(I) CAPM**	**(II) Carhart**	**(III) 6 Factor**	**(I) CAPM**	**(II) Carhart**	**(III) 6 Factor**
ESG	α	0.12888	0.08289	0.15919	0.51908	0.18311	0.22767	0.58637	0.66666*	0.77915*
diff.	MKT-Rf	0.11794*	0.14467*	0.09666*	-0.17108	-0.09647*	-0.105907*	0.37301*	0.19287*	0.14982
	SMB		0.91780*	0.84894*		0.67056*	0.656181*		0.38906*	0.34639*
	HML		-0.34460	-0.19769		-0.17690*	-0.255690*		-0.11340	-0.10316
	MOM		-0.19603	-0.14973*		0.19348*	0.201894*		-0.30454*	-0.30042*
	RMW			-0.04845			-0.230080			-0.028745
	CMA			-0.38668*			-0.007145			-0.05369
	Adj. R2	0.04344	*0.44890*	*0.45770*	*0.04681*	*0.47420*	*0.47560*	*0.14490*	*0.27960*	*0.27830*
ENV	α	0.03567	0.03099	0.07483	0.38291	0.07878	0.23519	0.05576	0.06288	0.02423
diff.	MKT-Rf	0.12163*	0.12333*	0.09720	0.09468	0.16686*	0.12667*	-0.08238	-0.12153	-0.10313
	SMB		0.093065*	0.89265*		0.58835*	0.54198*		0.12638	0.13303
	HML		-0.28111	-0.20489		0.09181	-0.05363		-0.04032	-0.7891
	MOM		-0.22986*	-0.20471*		0.10526	0.17501*	.	-0.05210	-0.05697
	RMW			-0.03276			-0.64226*			0.07361
	CMA			-0.21026			-0.22304			0.12023
	Adj. R2	*0.04015*	*0.40990*	*0.40500*	*0.00950*	*0.22500*	*0.27870*	*0.01043*	*0.00587*	*-0.004953*
SOC	α	0.30767	0.25056	0.31224	0.31888	0.16989	0.22012	0.10364	0.23509	0.43897
diff.	MKT-Rf	-0.8282	-0.06941	-0.10556*	-0.02794	-0.003318	-0.01291	0.33082*	0.20838*	0.14166
	SMB		0.69598*	0.64316*		0.307215*	0.29039*		0.06982	-0.03263
	HML		-0.14213	-0.03838		-0.094533	-0.20281*		0.20903	0.11766
	MOM		-0.09124	-0.05647		0.03206	0.03556		*-0.16779*	-0.17112*
	RMW			-0.04820			-0.28388			-0.59982*
	CMA			-0.29077			0.02137			0.22169
	Adj. R2	*0.02114*	*0.24330*	*0.24260*	*-0.005202*	*0.10010*	*0.11410*	*0.09802*	0.1434	*0.16640*
GOV	α	0.16914	0.05199	0.10244	0.04693	0.26748	0.18964	0.64800*	0.61903*	0.82901*
diff.	MKT-Rf	-0.02174	0.03456	0.02095	-0.11686*	-0.06752	-0.04180	0.11974	0.07328	-0.01670
	SMB		0.58339*	0.55725*		0.40359*	0.42334*		0.23875	0.18050
	HML		-0.25553*	-0.26019		-0.20927*	-0.023400*		-0.19570	-0.08287
	MOM		-0.03137	-0.01881		0.19050*	0.12322		-0.08823	-0.07133
	RMW			-0.09321			0.18584			-0.46934*
	CMA			-0.10765			0.27125			-0.37216
	Adj. R2	*-0.00528*	*0.19950*	*0.18940*	*0.02256*	*0.30800*	*0.31670*	*0.01490*	*0.03135*	*0.07656*

Appendix 5: Regression Results for the Equal-Weighted Portfolios

The table presents the risk-adjusted alphas (intercept), factor sensitivities for the risk factors size (SMB), value (HML), momentum (MOM), profitability (RMW) and investment (CMA), and the adjusted R^2 for the 36 equal-weighted portfolios based on the $30^{th}/70^{th}$ percentiles. For each investment segment, outputs of the three model specifications (I)-(III) are reported. **Panel A** reports results for the responsible portfolios, **panel B** reports results for the irresponsible portfolios, **panel C** reports results for the difference portfolios. The asterisk (*) indicates statistical significance at the 95% level.

Panel A

		EU			Japan			US		
		(I) CAPM	**(II) Carhart**	**(III) 6 Factor**	**(I) CAPM**	**(II) Carhart**	**(III) 6 Factor**	**(I) CAPM**	**(II) Carhart**	**(III) 6 Factor**
ESG	**α**	0.21078	0.33167*	0.266019*	0.39140*	0.21302*	0.18307*	0.17541	0.24088*	0.21481*
resp.	**MKT-Rf**	1.17158*	1.11467*	1.128434*	0.94978*	0.97300*	0.98177*	1.14850*	1.02843*	1.03479*
	SMB		-0.02007	0.009717		0.35753*	0.36578*		0.31572*	0.33365*
	HML		0.07822	0.10642		0.19721*	0.20526*		0.02404	0.05688
	MOM		-0.11732*	-0.129736*		-0.13297*	-0.15240*		-0.14977*	-0.14727*
	RMW			0.134577			0.09802			0.09189
	CMA			0.107875			0.07261			-0.08976
	Adj. R2	*0.09594*	*0.9638*	*0.9638*	*0.9031*	*0.947*	*0.9472*	*0.9194*	*0.9602*	*0.9605*
ENV	**α**	0.17809	0.27872*	0.221221	0.35466*	0.19922*	0.14189	0.11537	0.16236	0.144184
resp.	**MKT-Rf**	1.17164*	1.12312*	1.13917*	0.94251*	0.96514*	0.98278*	1.07917*	1.01039*	1.016810*
	SMB		-0.01122	0.01915		0.30774*	0.32304*		0.11772*	0.125804*
	HML		0.07216	0.07447		0.19525*	0.19454*		0.0327	0.036244
	MOM		-0.09675*	-0.11160*		-0.10838*	-0.15052*		-0.09502	-0.095173*
	RMW			0.10444			0.16729*			0.050168
	CMA			0.12707			0.16331*			-0.006419
	Adj. R2	*0.956*	*0.9589*	*0.9587*	*0.9011*	*0.9355*	*0.9385*	*0.9468*	*0.9602*	*0.9597*
SOC	**α**	0.25912*	0.384110*	0.34417*	0.35186*	0.21492*	0.19214*	0.14866	0.22479*	0.19087*
resp.	**MKT-Rf**	1.19097*	1.131009*	1.14796*	0.96608*	0.97786*	0.98415*	1.12573*	1.00278*	1.01807*
	SMB		-0.007127	0.02017		0.27982*	0.2862*		0.25239*	0.26012*
	HML		0.85263	0.05431		0.17109*	0.18422*		0.06117	0.03561
	MOM		-0.122505	-0.13859*		-0.15319*	-0.16581*		-0.14665*	-0.15011*
	RMW			0.05298			0.08337			0.07053
	CMA			0.1356			0.04462			0.08142
	Adj. R2	*0.9611*	*0.966*	*0.9659*	*0.91*	*0.9444*	*0.9441*	*0.932*	*0.9708*	*0.9712*
GOV	**α**	0.13864	0.26386*	0.23082	0.34547*	0.17366*	0.16969	0.17412	0.24376*	0.23022*
resp.	**MKT-Rf**	1.17102*	1.09546*	1.10421*	0.95114*	0.97287*	0.97511*	1.12995*	1.10385*	1.01834*
	SMB		0.11425	0.13119*		0.34380*	0.34426*		0.24933*	0.25600*
	HML		0.16700*	0.17094		0.21278*	9.19382*		0.05716	0.06264
	MOM		-0.12227*	-0.13033*		-0.14196*	-0.15068*		-0.13305*	-0.13289*
	RMW			0.06157			-0.01226			0.0394
	CMA			0.0692			0.03986			-0.01303
	Adj. R2	*0.9566*	*0.9662*	*0.9658*	*0.9043*	*0.9496*	*0.9491*	*0.9126*	*0.9454*	*0.9446*

Panel B

		EU			Japan			US		
		(I) CAPM	(II) Carhart	(III) 6 Factor	(I) CAPM	(II) Carhart	(III) 6 Factor	(I) CAPM	(II) Carhart	(III) 6 Factor
ESG irresp.	**α**	0.48989*	0.452086*	0.45524*	0.63332*	0.16831*	0.18349*	0.93243*	0.96356*	1.17776*
	MKT-Rf	1.11401*	1.103453*	1.10997*	0.84048*	0.92814*	0.92600*	1.26337*	1.08193*	1.01109*
	SMB		0.645397*	0.65164*		0.92106*	0.91553*		0.66867*	0.56268*
	HML		-0.001549	-0.04318		0.18819*	0.14136*		-0.08033	-0.16912
	MOM		-0.068015	-0.07457		-0.02035	-0.02362		-0.14602	-0.14881
	RMW			-0.03066			-0.10357			-0.62500*
	CMA			0.05339			0.02777			0.21201
	Adj. R2	*0.9146*	*0.9486*	*0.9479*	*0.7174*	*0.9495*	*0.9502*	*0.6076*	*0.6556*	*0.6658*
ENV irresp.	**α**	0.33708*	0.32555*	0.33784*	0.54015*	0.15633	0.162665	0.17228	0.2555	0.28172
	MKT-Rf	1.12250*	1.09786*	1.07812*	0.88377*	0.94927*	0.950203*	1.10774*	0.94446*	0.94244*
	SMB		0.63105*	0.60735*		0.76611*	0.767204*		0.42878*	0.40101*
	HML		0.02858	0.12122		0.18050*	0.184092*		0.09193	0.16229
	MOM		-0.08949*	-0.07006		-0.07583*	-0.077405*		-0.13828*	-0.14499*
	RMW			0.03544			0.015284			-0.12299
	CMA			-0.16033			0.004951			0.21388
	Adj. R2	*0.9233*	*0.9577*	*0.9577*	*0.7989*	*0.9527*	*0.952*	*0.8297*	*0.8909*	*0.8921*
SOC irresp.	**α**	0.40665*	0.36304*	0.36391*	0.54342*	0.16667	0.15685	-0.02386	0.07657	0.30039
	MKT-Rf	1.07477*	1.07284*	1.07541*	0.87492*	0.93326*	0.93854*	1.02501*	0.88490*	0.79575*
	SMB		0.62257*	0.62512*		0.75854*	0.75985*		0.30763*	0.23050*
	HML		-0.04134	-0.05897		0.17307*	0.13100*		0.19987*	0.25425*
	MOM		-0.06822	-0.07080		-0.11518*	-0.13528*		-0.09278*	-0.08124
	RMW			-0.01107			-0.02421			-0.54752*
	CMA			0.02105			0.09135			-0.20553
	Adj. R2	*0.9193*	*0.9536*	*0.9529*	*0.7951*	*0.946*	*0.9466*	*0.7713*	*0.8197*	*0.8449*
GOV irresp.	**α**	0.3899*	0.337815*	0.28636	0.96044*	0.26758*	0.26654*	1.03347*	1.03531*	1.21585*
	MKT-Rf	1.1221*	1.123915*	1.14627*	0.87534*	0.939*	0.93861*	1.229998*	1.10283*	1.05782*
	SMB		0.498762*	0.53448*		0.85718*	0.85793*		0.50246*	0.38045*
	HML		-0.008789	-0.05157		-0.11178*	0.13646*		-0.15484	-0.37277
	MOM		-0.027805	-0.04905			-0.10793*		-0.13309	-0.14950
	RMW			0.06651			0.02198			-0.62943*
	CMA			0.17891			-0.01972			0.5939
	Adj. R2	*0.9166*	*0.935*	*0.9348*	*0.7523*	*0.9349*	*0.934*	*0.5596*	*0.5809*	*0.5961*

Panel C

		EU			Japan			US		
		(I) CAPM	(II) Carhart	(III) 6 Factor	(I) CAPM	(II) Carhart	(III) 6 Factor	(I) CAPM	(II) Carhart	(III) 6 Factor
ESG	**α**	0.2333	0.073044	0.13909	0.19539	-0.093798	-0.05103	0.7093	0.675813	0.915670*
diff.	**MKT-Rf**	-0.05562	0.008491	-0.01458	-0.10566*	-0.041120	0.05158	0.1084	0.056524	-0.020137
	SMB		0.675658*	0.65399*		0.569084*	0.55609*		0.351530*	0.226913
	HML		-0.084075	-0.15602		-0.009606	0.05925		-0.104107	0.229517
	MOM		0.048244	0.053		0.110475	0.12652*		0.001769	-0.003911
	RMW			-0.16157			-0.18807*			-0.718510*
	CMA			-0.04516			-0.04609			0.312786
	Adj. R2	*0.017*	*0.3985*	*0.3944*	*0.044*	*0.6087*	*0.6169*	*0.004*	*0.01567*	*0.05911*
ENV	**α**	0.11319	0.0004576	0.0665	0.13896	-0.09197	-0.040671	0.009201	0.04627	0.09026
diff.	**MKT-Rf**	-0.4719	-0.0225363	-0.05716	-0.05510	-0.01213	-0.028392	0.031141	0.06290	-0.07080
	SMB		0.65254*	0.60026*		0.46393*	0.450515*		0.30964*	0.27309*
	HML		-0.0478886	0.04033		-0.01533	-0.005796		0.05948	-0.02347
	MOM		0.0061976	0.03937		0.03041	0.070862*		-0.04525	-0.05218
	RMW			-0.06534			-0.138488			0-17478
	CMA			0.027808*			-0.159601*			0.23133
	Adj. R2	*0.013*	*0.4262*	*0.4404*	*0.013*	*0.4695*	*0.4841*	*0*	*0.145*	*0.1607*
SOC	**α**	0.10172	-0.06845	-0.03039	0.14503	-0.097327	_0.08674	-0.22023	0.19509	0.06224
diff.	**MKT-Rf**	-0.11425*	-0.05544*	-0.06866*	-0.08753*	-0.040857	-0.04142	-0.09715*	-0.11486*	-0.21875*
	SMB		0.63988*	0.61702*		0.484277*	0.47988*		0.05383	-0.03174
	HML		-0.13271*	-0.11970		0.001395	0.04857		0.13895	0.21513*
	MOM		0.05322	0.06562		0.035866	0.02827		0.05189	0.0665
	RMW			-0.06038			-0.09406			-0.61968*
	CMA			0.10523			0.04549			-0.27592
	Adj. R2	*0.1013*	*0.4768*	*0.4718*	*0.034*	*0.4275*	*0.4271*	*0.026*	*0.0327*	*0.2011*
GOV	**α**	0.20546	0.02659	0.00541	0.29843	0.04483	0.0454	0.81165	0.744685	0.93835*
diff.	**MKT-Rf**	-0.04697	0.03118	0.045943	-0.07216	-0.03002	-0.03231	0.10361	0.090201	0.04305
	SMB		0.39470*	0.415351*		0.51894*	0.52001*		0.251716	0.12234
	HML		-0.18010*	-0.228926		-0.09192	-0.05271		-0.211744	0.43893*
	MOM		0.09340*	0.79118		0.02804	0.04049		-0.002021	0.01898
	RMW			0.008606			0.04776			-0.67046
	CMA			0.119035			-0.06082			0.61796
	Adj. R2	0.012	0.2241	*0.2159*	*0.02*	*0.5094*	*0.6069*	*0.001*	*0*	*0.0388*