Erasmus School of Economics
MSc Thesis Financial Economics

ESG on Stock Market Returns and Firm Fundamentals

Name student: Ozan Çelik
Student Id: 450814

Supervisor: Esad Smajlbegovic
Second assessor:

Date final version: 01 May 2021

Abstract

With an international data covering the 2010-2019 period, this research evaluates the relationship between environmental, social, and governance scores (ESG) and monthly stock returns. My initial results document a higher average monthly value weight returns for portfolios that carry lower E, S and G. Subsequently, I employ Fama and Macbeth (1973) regressions and find a positive risk premium for E and G scores for my full sample and a positive risk premium for S and G scores for my all-but micro sample. I then study the relationship between ESG and monthly returns for my US and European sub-samples separately. Consequently, for the US sample, my results show a significantly negative relationship between E as well as S scores and monthly returns. For the European sample the relationship changes sign and becomes positive. The governance score, on the other hand, appears to carry significant explanatory power when added to a CAPM model, and loses its significance when employed in more complex models. Next, I evaluate the predictive power of ESG scores on firm valuation using Tobin's q, price-to-earnings, and market to book ratios as valuation metrics. E and G scores carry a negative explanatory power on Tobin's q for the US and the European sub-samples. Moreover, the relationship between S score and Tobin's q is significantly negative for the European sub-sample, and insignificant for the US sub-sample. While all three E, S and G scores fails to predict price-to-earnings ratios, they have a significantly positive relationship with market to book equity ratios in both sub-samples.

Acknowledgements

I want to express my sincere gratitude to Dr Esad Smajlbegovic for his guidance. I would also like to thank Erasmus School of Economics for enabling me to come across with great minds, Erasmus Data Service Center team for their patience and assistance, and my family and those close to me for their continued support throughout this journey.

Table of Contents

1 INTRODUCTION

Sustainable investing continues to attract significant attention from academia, institutional as well as individual investors. The term sustainable investing refers to the process that combines Environmental, Social, and Governance (ESG) characteristics into the financial analysis (Global Sustainable Investment Alliance, 2018). According to the United Nations (UN) backed Principles for Responsible Investment (2020), sustainable investing can be favored, for example, to address ESG issues. Indeed, the relevance and the importance of these issues are growing each year. According to the World Meteorological Organization, each decade since the 1980s has been warmer than the previous one, marking 2011-2020 as the warmest ever on record (World Meteorological Organization, 2020). Climate change is causing an increased frequency of wildfires, increased sea levels along with increased atmospheric carbon levels. To address these matters, a growing number of money managers engage in shareholder activism to pursue board of directors to align their operations with a sustainable agenda.[1][2] And a growing number of corporations and sovereigns are pledging their commitments to sustainability.[3]

According to the Global Sustainable Investment Alliance (GSIA), Europe was the primary region in sustainable investing in 2018 with $14.08 trillion in assets, followed by the United States, and Japan with $12 trillion and $2.18 trillion in sustainable assets under management respectively. When compared with 2016, the increase in assets in Europe was a noteworthy 17%. For the US the increase was 38%, and for Japan, the increase was an astonishing 360%. On a global scale, as of the third quarter of 2020, the number of signatories acknowledging the Principles for Responsible Investment (PRI) increased by 29% on a year-on-year basis and reached to 3,300, representing $103.4 trillion of assets under management (Principles for Responsible Investment, 2020). Yet, the demand for sustainable investing is expected to accelerate further. While the share of European ESG mutual funds compared with the total European mutual funds stood at 15.1% in December 2019, the report from PWC highlights that this number will reach approximately 60% by the end of 2025 (PWC, 2020).

With such a significant shift to ESG, the motivating question is on how ESG affects firm financial performance. While academic results are inconclusive, according to the PRI, investing that includes at least one of the environmental, social and governance motives realize superior financial performance. Accordingly, with data covering the 2010-2019 period across 4 regions,

[1] See for example, Investors urge European companies to include climate risks in accounts.
[2] Also, BlackRock to Press Companies on Human Rights and Nature,
[3] For example, the Paris Agreement which is a legally binding international treaty signed by 191 countries in the world (United Nations, 2015).

I evaluate whether higher E, S, and G scores predict higher stock market returns. Although one can document a significant relationship between a variable of interest and stock market returns, the underlying reason may not be always connected to firm fundamentals (see for example Shiller (2003). Accordingly, I investigate whether environmental, social and governance scores predict various valuation metrics formed with firm fundamentals.

My analyses start with the formation of value weight and equal weight sort portfolios on environmental, social, and governance scores. Subsequently, I display the variation in average monthly returns for each underlying portfolio in excess of a matching portfolio created on size and book to market equity. My observations for average monthly excess returns from sort portfolios are somewhat divergent from PRI's claims. For portfolios created using the full sample and all-but micro firms, I find higher excess returns for lower E, S and G scored firms. More specifically, the difference between value weight returns on low E scored and high E scored portfolios formed with all-but micro firms reach to -1.28 percent per month and statistically significant at the 5 percent level.

Subsequently, with cross-sectional regressions on full sample, US as well as European sub samples, I test whether environmental, social and governance scores contain explanatory power over average monthly returns. I start this section by creating my own market, size, value, E, S, G score and profitability factors. For, the E, S and G factors I follow a worst (scored firm return) minus best (scored firm return) approach. For my full sample, the results show that environmental and social scores contain explanatory over monthly returns.

I, then, extend my research by employing these tests for my US and European sub-samples. The results from the US region are in line with those from the full sample. To offer a perspective, a one standard deviation increase in an environmental factor is expected to increase monthly returns 0.99 percent. In contrast, for the European region, the signs of the risk premiums on environmental and social scores turn to negative. The governance score, on the other hand, is significant only when employed for the full sample, for big and all-but micro size categories and has a positive risk premium.

Lastly, I test the predictive power of E, S and G scores over firm valuation metrics. Using Tobin's q as proxy I find that a E and G scores predict Tobin's q negatively in the US and the European sub-samples. However, the findings are not as robust as compared with those for the monthly stock returns. One unit increase in E and G scores is expected to result in a 0.002 and 0.001 percent decrease in Tobin's q ratio. Moreover, the social score contains explanatory power, only for the European sample. While significant at the 1 percent level, a unit increase in the social score translates to a 0.002 percent decrease in the Tobin's q of the European firms.

The rest of my paper is as follows. In section 2, I present the relevant literature and formulate my hypotheses. In section 3, I describe my research methodology. In section 4, I present my data gathering and my data transformation processes. Additionally, I show the descriptive statistics for my variables of interest and created factor variables. In section 5, I present results while in section 6, I discuss my limitations and conclude my research.

2 LITERATURE REVIEW AND HYPOTHESES

2.1 Relevant Literature on Sustainable Investing

Over the past decade, the number of academic articles on sustainable finance has seen a significant surge and the majority shows a substantial effect of sustainability measures on financial metrics. For example, Delis, de Greiff, Iosifidi and Ongena (2019) investigate how bank loan spreads are affected by firm exposure to climate risk. Their findings point out a positive relationship between an increase in firm fossil fuel reserves and firm cost of debt, especially after 2015 when the Paris Agreement was signed.

For the US private equity market Barber, Morse, and Yasuda (2021) finds a negative relationship between ESG and fund returns. The authors analyze the sustainable investor behavior by comparing the willingness-to-pay for traditional venture capital (VC) fund investors and impact VC fund investors. Subsequently, they report that impact VCs earn as low as 4.7% annually compared to their traditional peers. Yet, investors are willing to pay 3.4% to 6.2% more for impact VCs due to the non-monetary utility they receive from investing in those impact funds (Barber, Morse, & Yasuda, 2021).

Moreover, Baker, Bergstresser, Serafeim, and Wurgler (2018) analyze the US debt market and find a similar result. They show that green municipal bonds contain an issuance premium when compared with ordinary bonds of similar nature. Similarly, they argue that some significant part of investors is willing to pay a premium since they receive non-pecuniary benefits from holding green bonds.

Bernstein, Gustafson and Lewis (2019), instead, investigate whether real estate prices in the US reflect flood risks attributed to prospective rising sea levels. Analyzing sales of more than 460 thousand real estate properties, their research uncovers an average 6.6 percent discount for properties exposed to rising sea level risk compared with unexposed properties of the same feature. This indicates that the US real estate market prices already incorporate an aspect of environmental risk.

In his paper Krüger (2015) evaluates how shareholders react to developments in firm corporate social responsibility (CSR) while they measure shareholder reaction with cumulative abnormal returns. Krüger (2015) indicates that shareholders react negatively to CSR news when the underlying news signals an agency problem inside the firm, as argued by Benabou and Tirole (2010). On the other hand, the shareholder reaction is positive if the CRS news demonstrates an improvement towards an already known social irresponsibility (Krüger, 2015).

Moreover, covering the 1976-2006 period Hong and Kacperczyk (2009) show that sin stocks have less institutional ownership, receive less analyst coverage, and earn higher returns than comparable stocks. Sin stocks refer to companies engaged in activities considered unethical or immoral by the general public such as alcohol, tobacco, and gambling. According to Hong and Kacperczyk (2009), these results are derived from the litigation costs carried by sin stocks and investors constrained by social norms.

Some articles, on the other hand, document a mixed relationship between stock returns and ESG. Pastor, Stambaugh, and Taylor (2019) create a model that considers ESG as an investment criterion. They argue that differences in ESG taste are a must for the ESG industry to exist and investors indeed perceive different tastes from holding green stocks. They show that, in equilibrium, brown stocks outperform green stocks due to their high climate betas while climate betas measure firms' exposure to climate risk. On the other hand, they argue that green stocks outperform when then the ESG factor is hit by a positive shock.

Likewise, Pedersen, Fitzgibbons, and Pomorski (2020) evaluate the relationship between the E, S, and G scores as well as the overall ESG scores of firms and their subsequent financial performance. Their results show a negative relation between E, S, and overall ESG scores and expected returns. On the contrary, they document a positive relationship with the governance score and expected returns. Lastly, Pedersen, Fitzgibbons, and Pomorski (2020) show that ESG screened portfolios can carry inferior ESG scores than non ESG screened portfolios. They advocate that non-screened portfolios are able to take short positions in low ESG scored firms to finance long positions in higher ESG scored firms. However, for ESG screened portfolios, firms with inferior ESG scores do not appear in the investment universe, making it infeasible for portfolio managers to take advantage of the situation.

And how do fund managers incorporate ESG into their investment management process? To answer this question Duuren, Plantiga, and Scholtens (2016) employ a survey to fund managers domiciled in the US, UK and continental Europe. Their first result is that ESG investing results in low turnover. Indeed their survey shows that in 2012, 51% of the sample investment managers did not sell a single stock due to a positive ESG signal and 39% did not buy a single stock because of a negative signal. Moreover, Duuren, Plantiga, and Scholtens (2016) illustrate that US fund managers are rather skeptical about the benefits of ESG investing while European fund managers are optimistic. Their findings also highlight that fund managers analyze all E, S and G scores separetely. Lastly, the proportion of fund managers following the exclusionary screening approach in portfolio construction was 20% in Duuren, Plantiga, and Scholten's (2016) sample.

For the first part of my research, my focus is on the relationship between ESG and stock market returns. The latter section showed a rather inconclusive result for the relationship between ESG and stock market returns. According to Giese, Lee, Melas, Nagy, and Nishikawa (2019), this is due to differences in research methodology and the use of various ESG datasets. Yet, there is evidence that portfolio managers investigate E, S, and G score components separately. And portfolio managers in the US and Europe perceive the effects of integrating ESG into their portfolios differently. Taking these into consideration, I formulate the following hypotheses.

H1: A lower environmental score predicts higher average monthly returns for the full sample

H2: A lower social score predicts higher average monthly returns for the full sample

H3: A lower governance score predicts higher average monthly returns for the full sample

H4: E, S and G scores predict stock market returns negatively in the US sub-sample and positively in the European sub-sample.

2.2 ESG and Firm Fundamentals

Research on ESG and firm fundamentals is rather limited when compared to research that considers ESG and firm financial performance. Ponkratz, Bauer, and Dervall (2019), investigate the relationship between firm performance and firm heat exposure. They define a day as hot if the temperature exceeds 30 degrees Celsius as well as if it exceeds the 90^{th} percentile location and season average. Using data from 93 countries, they find a significantly negative relationship between firm revenues, operating income, and an additional day of heat exposure.

Moreover, in their paper, Aouadi and Marsad (2018) use ESG controversies in more than 4300 international firms to investigate the relationship between this indicator and firm value. Using Tobin's q as a proxy for firm value, Aouadi and Marsad (2018) document a positive relationship between corporate social performance and the value of high attention firms. These firms are located in countries with greater press freedom and receive more attention on internet searches. Likewise, in their paper Surroca, Aguilera, Desender, and Tribo (2020) investigate the relationship between CSR and shareholder value using international data, consisting of countries covering 81% of world GDP. Surroca et. al. (2020) too proxy firm value with Tobin's q and find a positive relationship between CSR and shareholder value in liberal market economies. Conversely, for coordinated market economies they document an opposite relationship.

Hong and Kacperczyk (2009) test, amongst others, the variation within certain valuation variables between sin stocks and non sin stocks. Correspondingly they find a lower market to book, price to earnings, and price to EBITDA (earnings before interest, tax, depreciation and amortization) ratios for sin stocks.

Furthermore, Pedersen, Fitzgibbons, and Pomorski (2020) analyze the predictive power of environmental, social, and governance scores over future profitability, where they use gross profits over assets as a proxy for profitability (see Novy-Marx (2013)). Accordingly, the authors find that the governance score predicts profitability positively, and the relationship is highly significant and economically large in magnitude. Additionally, Pedersen, Fitzgibbons, and Pomorski (2020) asses the explanatory power of ESG on firm valuation using the ratio of price to book equity (P/B) as the valuation metric. Correspondingly, they find a highly significant and negative relationship between the governance score and P/B.

Cao, Titman, Zhan, Zhang (2018) display an underpricing for low ESG firms and an overpricing for high ESG firms. They connect their results with a mispricing story in the equity markets. Typically, institutional investors with buying or selling mispriced securities would drive prices back to their fair values. However, with the shift towards socially responsible investing, money managers stay away from low ESG stocks causing the mispricing to persist.

Using a dataset covering FTSE 350 between 2004 and 2013 Li, Gong, Zhang, and Koh (2018) find a positive relationship between ESG disclosure and Tobin's q, as well as ESG disclosure and return on assets. Subsequently, Li et. al. (2018) investigate the relationship between these two variables of interest, this time, with environmental and social disclosure.

Accordingly, they document a relationship that is consistent with their findings considering total ESG disclosure.

The second part of my research focuses inspecting the link between E, S and G and firm fundamentals. Section 2.2, presented a positive relationship between firm corporate social performance and firm valuation metrics. Moreover, the evidence showed a positive link between better firm governance and firm profitability as well as firm value. Correspondingly I make the following hypothesis:

H5: A higher E, S or G score predicts a higher firm value where firm value is proxied with Tobin's q.

Popularized by Tobin (1969), Tobin's q shows how well a firm makes use of its resources and it is commonly used as a proxy for firm value (see, for example, Morck, Shleifer, and Vishny (1988), and Bergstresser, and Philippon (2006)). The exact procedures on how I construct Tobin's q in this research can be found in the ESG and Firm Valuation part of the methodology section.

2.3 Asset Pricing

The prominent Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) assumes a linear relationship between a security's expected return and its exposure to systematic risk. Under the CAPM paradigm, the residual risk is diversifiable, thus, does not lead to higher expected returns. As a result, an investor can achieve higher expected returns only by increasing its exposure to systematic risk. In the CAPM model, the systematic risk is denoted by the Beta of a security which is the covariance between the return of a firm and the market, both in excess of the risk-free ratio. In line with CAPM, the efficient market hypothesis (EMH) rests on the assumption that at any time, the price of a security fully reflects all available information (Fama, 1970). Consequently, under EHM, any departure from the expected return of a security can be explained by a piece of new information about the fundamental value of that underlying security (Shiller, 2003). On the other hand, any price deviation from a firm's fundamental value would lead to a correction by arbitrageurs. Ultimately, under the strong as well the semi-strong form of the EHM, it would be impossible for investors to produce consistent anomalous returns.

According to Fama and French (1992), although Beta was able to explain average stock returns in the 1926-1968 period, its explanatory power was starting to diminish. Similarly, the research was starting to provide evidence for patterns in average returns that were not explained by the risk-based explanation of the CAPM. Among the most eminent was the positive relationship between the book equity to market equity ratio (BE/ME) and the average returns of US stocks documented by Rosenberg, Reid, and Lanstein (1985). Correspondingly, Banz (1981) showed a negative relation between firm size and average returns of US stocks. In their renowned article, Fama and French (1992) displayed size, and the book-to-market equity, when included with the CAPM's, did a remarkable job explaining average stock returns.

Asset market anomalies and asset pricing models are not limited to those described above. In their research, De Bondt and Thaler (1985) show a reversal effect amongst the winning stocks and the losing stocks of the past 3 to 5 years. Put differently, their portfolio consisting of past losers outperforms the market approximately 3 years after the portfolio construction while their portfolio of past winners underperforms it. Subsequently, for a shorter portfolio horizon, Jagadeesh and Titman (1993) investigate whether it is possible to attain significant returns considering a portfolio of past winners or losers. They find that by constructing a long-short portfolio from the winners of past 3-to-12-months and the losers of the same time frame, one can attain significantly positive returns in the upcoming 3-to-12 months.

Several years later, Carhart (1997) extended the Fama and French three factor model (1993) by adding a momentum factor. While his main interest was on evaluating the mutual fund performance over consecutive periods of time, his model took a grand place in asset pricing and it is used widely as a benchmark. Nevertheless, the explanations on the reversal effect as well as the momentum effect do not show a direct relationship with a change in firm fundamentals and leans more to a behavioral based explanation. Indeed, both de Bondt and Thaler (1985) and Jagadeesh and Titman (1993) connect their findings with the overreaction of the market participants.

Another prominent asset market anomaly, namely gross profitability, appears in Novy-Marx (2013). He defines gross profitability as the ratio of gross profits (revenues deducted from the cost of goods sold) to assets. With his spanning from 1962 to 2010, Novy-Marx (2013) that more profitable firms earn significantly higher returns. His results cannot be explained by the three-factor model of Fama and French, since highly profitable firms his sample are growth firms- firms with low BE/ME ratios- and are larger in size. Additionally, the gross profitability in Novy-Marx (2013) did a superior job in explaining monthly return than the value factor.

3 METHODOLOGY

3.1 Return on Portfolios sorted on E, S and G Scores

Similar to Fama and French (2008), I follow a double sort approach to assess the relationship between stock returns and environmental, social and governance scores. At the end of each year, I create three equal-weight and value-weight sorts on size and variable characteristics. I use the 50th percentile market capitalization and above to differentiate firms as big caps. Micro caps, although hold a tiny place in terms of total sample market capitalization, can be plentiful in number, driving equal-weighted returns significantly. In order to investigate their effects too, I use the 20th percentile of market capitalization to differentiate between micro and small caps. For the variables of interest, I sort portfolios according to the 30th and the 70th percentiles and use all but micro stocks to determine the sort breakpoints. I call portfolios below the 30th low, above the 70th percentile high, and in between medium. Although not mutually exclusive, I create low, medium and high sort portfolios also for the total sample and all but micro size groups. All in all, I form 15 equal and value weight portfolios using the E, S and G score breakpoints. Simultaneously, I create matching portfolios formed on size and book to market equity. Finally, I calculate the average monthly returns for E, S and G portfolios in excess of matching size and book to market equity portfolios and call these returns excess returns.

3.2 ESG on Predicting Stock Market Returns

According to Fama and French (2008), while the sorts approach offers an outlook on how performance differs across portfolios created on various size and variable characteristics, it does not provide a clear picture on which variable has explanatory power on average returns. Since, my data consists of both time series and cross-sectional observations, a solution for this would be to employ Fama-Macbeth (FM) regressions (1973). FM regressions are used commonly in asset pricing to evaluate the relationship between risk factors and monthly excess returns. While simple OLS models assume independent error terms for the cross-sectional tests, FM regressions are an effective tool to correct for this bias in the error terms. Then, I start the second part of my analysis by constructing factor portfolios on the underlying variables. In the section below, initially the factor creation is explained, following the steps of FM regressions are shown.

Initially, I construct factor portfolios for the E, S and G scores using a low minus high approach and call these factors WMBE, WMBS and WMBG where WMB denotes worst minus best. Each year, I calculate the small and big size breakpoints from median sample market capitalization. Subsequently, I determine the low and high breakpoints using the 30th and the 70th environmental, social and governance score percentiles. Thereafter, I form 4 mimicking portfolios for each variable of interest. These portfolios are small low, small high, big low and big high, (S/L, SH, BL, BH). Last but not least, I. take the difference between the average low and high portfolios. This process is illustrated below:

$$WMB_{E,S,G} = \frac{1}{2}(Small\ Low + Big\ Low) - \frac{1}{2}(Small\ High + Big\ High)$$

Then, I include the created ESG risk factors into some of the most commonly used asset pricing models, also discussed briefly above in section 2, to infer the relationship with monthly returns. The first model is the Capital Asset Pricing Model of Sharpe (1964) and Lintner (1965), which is illustrated below as:

$$R_{it} - R_{ft} = \alpha_i + \beta_{RMRF}[RMRF] + \epsilon_{it} \quad (1)$$

where R_{it} is the return on firm i at time t, R_{ft} is the risk-free rate at time t, RMRF is the market return in excess of risk-free rate and ϵ_{it} is the error term.

Second model is the Fama and French (1993) (FF) 3 factor model, which is illustrated below as:

$$R_{it} - R_{ft} = \alpha_i + \beta_{RMRF}(RMRF) + \beta_{SMB}(SMB_t) + \beta_{HML}(HML_t) + \epsilon_{it} \quad (2)$$

where SMB (small minus big) and HML (high minus low) are the returns on mimicking portfolios created on size and book to market equity. Calculation of these portfolios are discussed below.

As a robustness check a third model is considered. It is the FF 3 factor model + Profitability and is described below.

$$R_{it} - R_{ft} = \alpha_i + \beta_{RMRF}(RMRF) + \beta_{SMB}(SMB_t) + \beta_{HML}(HML_t) + \beta_{PMU}(PMU) + \epsilon_{it} \quad (3)$$

where PMU (profitable minus unprofitable) is the profitability factor of Novy-Marx (2013). Profitability is measured as the ratio of gross profits, which is the difference between revenues and cost of goods sold (COGS), divided by total assets.

In order to employ the three models mentioned above, I create my own market (RMRF), size (SMB), value (HML) and profitability (PMU) factors. RMRF is simply the average monthly value weighted returns in excess of the region-specific risk-free rates. For SMB, I first calculate the small value, big value, small neutral, big neutral, small growth and big growth breakpoints for each year end. The 30th and the 70th book to market equity percentiles define value, neutral and growth while the median sample capitalization percentiles define small and big size groups. Then the difference between the average small and average big portfolios is taken. HML and PMU are created similarly to WMB. Each year, small and big breakpoints are created using the median sample market capitalization. Likewise, low and high weight portfolios on book to market equity and profitability are created using the 30th and the 70th variable of interest percentiles. Then the difference between the average high weight portfolios is taken against the average of low weight portfolios. The process for creating the size, value and profitability factor can be summarized below as the following:

$$SMB = \frac{1}{3}(\text{Small Value} + \text{Small Neutral} + \text{Small Growth}) - \frac{1}{3}(\text{Big Value} + \text{Big Neutral} + \text{Big Growth})$$

$$HML = \frac{1}{2}(Small\ High + Big\ High) - \frac{1}{2}(Small\ Low + Big\ Low)$$

$$PMU = \frac{1}{2}(Small\ Profitable + Big\ Profitable) - \frac{1}{2}(Small\ Unprofitable + Big\ Unprofitable)$$

Returning to the FM regressions, the procedure consists of two-steps. In the first step, time-series regressions are run on monthly excess returns to find the factor loadings, which can be seen below as;

$$R_{it} - R_{ft} = a_{it} + \beta_{1it}RMRF + \beta_{2it}SMB + \beta_{3it}HML + \beta_{4,5,6it}WMB_{E,S,G} + \beta_{7it}PMU + \epsilon_{it} \qquad (4)$$

where $R_{it} - R_{ft}$ is the monthly return for firm (i) in excess of risk-free rate at time t. Moreover, the βs are the estimated factor loadings in the first stage Fama-MacBeth regressions.

In the second step, cross-sectional regressions on the estimated factor loadings are run separately for each time period.

$$R_{it} - R_{ft} = \hat{\alpha}_{it} + \Upsilon_{1t}\,\hat{\beta}RMRF_i + \Upsilon_{2t}\,\hat{\beta}SMB_i + \ldots + \Upsilon_{6t}\,\hat{\beta}WMBG_i + \Upsilon_{7t}\hat{\beta}PMU_i + \epsilon_{it} \qquad (5)$$

where ϒs are the so called factors risk premiums on the $\hat{\beta}$s estimated in the first stage (4) Fama-Macbeth regressions, ϵ_{it} is the error term and intercept $\hat{a}_{it}$ is the unexplained part of the monthly returns. Lastly, risk premiums are averaged over the sample observation period and t-statistics are created:

$$t - statistic_i = \frac{\sum_{i=1}^{7} \hat{\Upsilon}_i}{\sqrt{T}}$$

where $\hat{\Upsilon}_i$ is the average of the estimated risk factor in equation (5).

3.3 ESG and Firm Valuation

For the third part of my analysis, I am interested in the explanatory power of E, S and G scores over firm value based on firm fundamentals. For this part of my research, I follow panel regressions where I decide between the two most commonly used approaches, month fixed effects and random effects by employing a commonly used Hausman (1978) test. Random effects assume the error term is independent of the regressors while fixed effects allow this kind of an association. Accordingly, if one fails to reject the null hypothesis of the Hausman test then the random effects are to be followed. On the other hand, the rejection of the null hypothesis favors the fixed effects model. Similar to Fama-Macbeth regressions, in panel regressions, one can account for the correction of the standard errors, by attaining robust standard errors or cluster standard errors for a specified variable.

After deciding, which model to employ, I decide on fundamental variables to be included in panel regressions. My main variable of interest is Tobin's q which is defined as the ratio of the market value of a firm over the replacement cost of its assets. While the exact formulation Tobin's q differs amongst research, Chung and Pruitt (1994), offers a relatively simple yet powerful approximation. Their model can see below as:

$$q = \frac{(MVE + PS + DEBT)}{TA}$$

Where MVE is market capitalization, PS is the liquidating value of preferred stock, Debt is the sum of short-term debt and long-term debt in excess of short-term assets, and TA is the book value of total assets. A q measure of one is considered as the point where the market value of a firm is equal to its intrinsic value. Hence, a q measure above (below) one indicates that the firm is overvalued (undervalued).

I choose to follow this estimation since it contains a limited number of variables which reduces my exposure to possible missing cases. Additionally, according to Chung and Pruitt (1994), their model is able to capture at least 97 percent of variability in a Lindenberg and Ross (1981) model, which employs a more conventional and extensive Tobin's q measure.

3.4 Multicollinearity, Heteroskedasticity and Autocorrelation Checks

Multicollinearity might cause an issue when employing the models discussed above. Accordingly, variance inflation factors for each model on each size group are calculated where a variance inflation factor of 10 and above can be interpreted as a signal of multicollinearity. Due to its nature, the data is expected to be heteroskedastic. Moreover, the residuals are likely to be autocorrelated up to a certain lag. To assess the autocorrelations, a pooled OLS regression, containing model 3 is run for the total sample. Then, the Breusch-Godfrey test is employed on the model to determine whether autocorrelation possesses an issue. A solution for heteroskedasticity and autocorrelation would be to employ Newey West standard errors for the FM regressions. For the fixed and random effects robust as well as clustered standard errors are also useful in dealing with autocorrelation issues.

4 DATA

4.1 Data Gathering

My primary regions of interest are the US and Europe, but I also include the Asia-Pacific and Latin American regions in my research. For each region, I choose the largest stock market indices by market capitalization. For the US this is the S&P 500 index. For Europe, this is the Stoxx Europe 600 index. Similarly, for Asia – Pacific I follow the same principle, however, including only the developed regions. Accordingly, the largest indices are from Australia, New Zealand, Hong Kong, and Japan. Correspondingly I gather the constituents of the Australian All Ordinary Index, Nikkei 225, Hang Seng, New Zealand Stock Exchange 50. For Latin America, the countries with the largest stock exchanges are Brazil and Mexico. The chosen indices for these countries are the indices are Brazilian Bovespa and IPC–Mexico indices. Then, I gather index constituent lists from Capital IQ offered in Compustat within Wharton Research Data Services. All of the stock market indices in this research are constructed with the largest firms in their local markets. The exception is the Australian All Ordinary Index. The reason behind the decision to include this index was solely due to the coverage of Capital IQ. All in all, the majority of my sample consists of large, liquid stocks with relatively high analyst coverage. Besides, my research period is between 2010 and 2019. Post 2010 period was specifically determined. This is because the quality of ESG data and the quantity of firms publishing ESG information increases considerably, especially after the financial crisis, due to numerous laws and restrictions presented by regulatory authorities.

Subsequently, I gather annually and quarterly US fundamentals data from Compustat and monthly US stock price data from the Center for Security Research in Prices (CRSP). For non-US firms, I gather fundamentals data and stock price data from the Eikon database offered within Thomson Reuters. Likewise, I gather environmental, social and governance score data from the Asset4 database within Refinitiv. Asset4 is a commonly used database in research (see for example Aouadi and Marsad (2018), and Surroca et. al. (2020)). Asset4 assesses a firms' ESG performance over 450 criteria points and covers more than 80 percent of total market capitalization (Refinitiv, 2021).

Regarding firm fundamentals, I gather the number of shares outstanding to calculate market capitalization. I gather revenues, cost of goods sold and total assets to calculate firm profitability. Moreover, I gather book equity to form book equity to market equity ratio and MSCI's Global Industry Classification Standard (GICS) codes to identify industries and sectors. For stock level information, I export holding period return from CRSP, total return index from Reuters and stock prices from both of the databases. Holding period return and total return index account for firm specific events such as stock splits as well as dividends. Correspondingly, I use these variables as my return variable. Lastly, I gather monthly risk-free rates from Kenneth French's data library for the US, Europe, Asia – Pacific excluding Japan, Japan, and Emerging Markets regions. Accordingly, I match the region-specific risk-free rates with my own regions and risk-adjusted returns.

4.2 Data Transformation

All the fundamental variables and stock prices are denoted in their local currencies. Correspondingly, I convert locally denominated currencies into US dollars using end of year local currency/USD rates. Additionally, firms with observations less than 2 full years, that is from January (t) to December (t+1), firms with a maximum price of less than 1 dollar during their total observation period and financial firms (GICS sector code 40) are excluded from my research.

Subsequently, each year size breakpoints are calculated from the 20^{th} and the 50^{th} percentiles. Micro caps are those below the 20^{th} percentile while big caps are those larger than the median. In between, are the small caps. Additionally, 4 regions, Asia-Pacific (AP), Developed Europe (DE), the US, and Emerging Markets (EM) are created. EM consists of firms domiciled in Poland, Czechia, Singapore, Indonesia, and stocks from JSE/FTSE 40, Brazilian Bovespa, and IPC– Mexico indices. To present an overview, AP covers 31% of the sample firms, accounting for 17% of sample market capitalization. Most notable, the number of US firms makes up 28% of the total sample while covering 48% percent of the total sample market capitalization. For DE and EM, the number of firms forms up 34% and 8% of the sample while these regions represent 6% and 29% of the sample size respectively.

4.3 Descriptive Statistics and Correlation Coefficients

Table 1 reports the descriptive statistics for size, book to market equity, Tobin's q, profitability, E, S, and G scores for the sample, micro, small, big as well as all but micro firms. Equal and value weight returns are in percentage points and in excess of region-specific risk-free rates. The number of micro-cap firms compared with their relative market show a big difference. While making up 23 percent of all sample firms, micro caps account only for 0.94 percent of the sample market capitalization. All in all, sample micro caps carry lower E, S, G scores, have lower equal and value-weight returns, and book to market equity ratios. Yet, they appear to be more profitable. It is important to note that 48.91 percent of all micro caps belong to the AP region. While almost all the indices gathered for this research consists of the largest firms in their local stock exchanges, Australian All Ordinary Index includes all publicly traded Australian stocks. Unsurprisingly, Australia alone accounts for one third of all sample micro caps. On the other hand, the share observations attributed to the US firms is 23.22 percent while the US firms account for 48.22 percent of the total sample market cap. Each region and countries' share in the number of observations, micro, small and big caps as well in total sample market capitalization can be found in Table A1.

Returning to Table 1, one can notice that value-weight returns have smaller standard deviations compared with the equal-weight returns. Both equal weighted and value weighted returns appear to be relatively stable across all size groups except for the equal weighted micro caps which have monthly returns 67 basis points lower than their small cap peers. The lowest book equity to market equity, E, S, and G score values are also observed for micro caps. On the other hand, micro caps appear to be more profitable and have lower valuation. Indeed, the smallest mean for Tobin's q is observed for this size group.

Both Tobin's q and book to market equity have heavy tails which raise concerns for the feasibility of the statistical models. Before forming the models, I normalize these variables by taking their natural logarithms. At the same time, this means excluding negative observations for Tobin's and Book to Market from statistical models. The negative Tobin's q observations only make up 0.09 percent of all the Tobin's q observations in the US and the European sub samples while the same proportion reaches up to 2.78 percent for negative BME observations. Indeed, the kurtosis of the sample Tobin's q and book to market equity this value trail down to 5.66 and 3.92 respectively. E, S, and G sores appear to be normally distributed while skewed slightly to the left.

Additionally, to give an overview of how variables differ amongst sectors one can look at Table A2 which displays the number of firms as well as the averages of the E, S, G scores, profitability, book to market equity and Tobin's q across 10 GICS sectors for the two regions of interest. The table also highlights the proportion of the sector size compared with the total size of the two sub samples. The real estate sector, while carrying the lowest environmental, social, and governance scores with mean values of 41.88, 52.92, and 49.28 respectively, also has the lowest proportion in the total US and the UK sub sample size. Conversely utilities, while having the least number of firms, contain the highest environmental and governance scores. Laslty, the highest social score is seen in the materials sector.

Table 1
Descriptive Statistics

Environmental, Social and Governance scores are gathered from Refintiv's Asset4 database. Tobin's q is a proxy for firm value. Profitability is gross profits (Revenues – COGS) divided by total assets. BME is book equity divided by market equity. Market capitalization is in millions of USD. Size breakpoints are the 20th and 50th market cap percentiles. Returns are in excess of region-specific risk-free rates.

	Number of Firms	Percent of Total Market Cap	**Market Cap**		**EW Average Monthly Returns**		**VW Average Monthly Returns**	
			Mean	Std Dev	Mean	Std dev	Mean	Std dev
Sample	3,919	100	16,738	39,067	1.04	6.41	1.07	3.13
Micro	888	0.94	795	2,436	0.44	13.71	1.06	3.12
Small	1,243	6.49	3,641	1,410	1.11	9.36	1.08	3.09
Big	1,788	92.51	31,616	51,747	1.12	8.26	1.08	3.14
All-but Micro	3,031	99.06	20,766	42,777	1.13	6.98	1.08	3.13

	Env. Score		**Social Score**		**Gov. Score**	
	Mean	Std Dev	Mean	Std dev	Mean	Std dev
Sample	46.47	28.72	52.02	23.79	52.81	21.91
Micro	25.24	25.16	35.40	20.72	44.61	20.34
Small	43.13	26.18	48.65	21.76	49.85	20.99
Big	56.89	26.30	60.62	22.04	57.84	21.74
All-but Micro	51.70	27.09	56.10	22.69	54.82	21.81

	Env. Score		**Social Score**		**Gov. Score**	
	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
Sample	-0.19	1.81	-0.10	2.05	-0.09	2.10
Micro	0.76	2.40	0.57	2.62	0.19	2.25
Small	-0.05	1.94	0.00	2.17	-0.01	2.12
Big	-0.65	2.47	-0.44	2.42	-0.31	2.21
All-but Micro	-0.39	2.09	-0.25	2.20	-0.18	2.13

	Tobin's q		**BME**		**Profitability**	
	Mean	Std dev	Mean	Std dev	Mean	Std dev
Sample	1.90	3.07	14.40	44.37	0.66	5.10
Micro	1.31	1.44	10.87	48.11	0.99	11.45
Small	1.91	4.49	18.78	49.97	0.69	0.64
Big	1.98	2.20	13.11	38.43	0.53	0.66
All-but Micro	1.96	3.19	15.29	43.33	0.59	0.66

	Tobin's q		**BME**		**Profitability**	
	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
Sample	14.66	320.34	3.57	24.71	125.80	1,613.43
Micro	5.95	56.77	3.08	37.59	56.75	324.02
Small	4.49	212.51	3.39	16.21	3.04	21.77
Big	7.42	97.00	3.95	21.75	14.97	452.13
All-but Micro	14.38	303.37	3.75	19.65	10.74	299.54

Correlation coefficients are presented in Table 2. E, S, and G scores are positively correlated with each other as well as with firm size. Profitability and BME appear to be almost non-correlated. On the other hand, significant differences across regions are observed for correlation coefficients. These differences can be seen in Table A3 which shows correlation matrices across regions. For the US, a renowned scenario between profitability and book to market ratio is observed. These variables, unlike for the total sample, are negatively correlated while the correlation coefficient stands at -0.28. Likewise, a positive correlation between EM and AP firms are reported. For the US and the DE, the coefficient changes sign.

Table 2
Cross Sectional Correlations

Correlation coefficients are reported below. Profitability is gross profits (Revenues-COGS) divided by total assets, BME is the ratio book equity to market equity.

	MC	Environ-mental Score	Social Score	Governance Score	Profitability	BME	Tobin's Q
Market Cap	1.000						
Environmental Score	0.265	1.000					
Social Score	0.286	0.731	1.000				
Governance Score	0.244	0.476	0.563	1.000			
Profitability	-0.016	-0.008	-0.014	-0.004	1.000		
BME	-0.044	0.143	-0.076	0.009	0.001	1.000	
Tobin's Q	0.089	-0.108	-0.049	0.009	-0.020	-0.018	1.000

4.4 Sort Breakpoints and Factor Variables

At the end of each year, sort breakpoints are created on size, environmental, social and governance scores as well as on firm profitability and book to market equity. Initially, size is split between micro, small, and big using the 20th and the 50th percentiles. For the variables of interest, low, medium, and high breakpoints are determined from the 30th and 70th deciles, using all but micro firms. Firms are then assigned to the relative size and variable of interest category, therefore, 9 portfolios for each variable of interest are created each year.

Yearly sort breakpoints can be found in Table A5. Size breakpoints drop by $276 million, and $474 million for micro and big firms at the beginning of the sample. They then stay relatively stable for micro caps while increasing steadily for big caps only to show a leap of $350 million and $1.34 billion for the micro and the big size groups in 2017. For E, S and G breakpoints, a constant increase is observed while for profitability breakpoints a constant decrease is observed. Book to market equity and size breakpoints appear to move coherently.

Moreover, factors on market return excess return, size, value, profitability, E, S, and G were created using all but micro firms. These factors were created according to the methodologies described in the previous sections, using the year-end breakpoints. Similar to sort portfolio creation, the 30th and the 70th percentiles for variables of interest and the median for size was determined prior to portfolio formation. Table 3 displays the descriptive statistics as well as the correlation coefficients for these factors. While SMB and HML are skewed to the right, the remaining factors are skewed to the left. Moreover, kurtosis values indicate fat tails. The majority of the factors appear to be highly correlated with each other. Especially, PMU variable is highly correlated with 5 of the sample factors. Governance factor on the other hand, appear to be the outlier. Compared with the other factors, the governance factor has relatively lower correlation coefficients.

Table 3
Factor Variable Descriptives and Correlation Coefficients

RMRF, SMB (small minus big) and HML (high minus low) are the Market, Size and Value factors created using sample data. WMBE, WMBS, and WMBG are the worst minus best factors formed on environmental, social and governance scores. PMU (profitable minus unprofitable) is the profitability factor where profitability is measured as gross profit (Revenues – COGS) divided by total assets.

	Mean	SD	Skew	Kurt		RM -RF	SMB	HML	WMBE	WMBS	WMBG	PMU
RMRF	-0.01	0.04	-0.93	3.20	RM -RF	1.00						
SMB	0.01	0.02	0.82	4.02	SMB	-0.45	1.00					
HML	0.01	0.03	0.37	3.96	HML	-0.59	0.73	1.00				
WMBE	-0.02	0.03	-0.49	4.31	WMBE	0.55	-0.75	-0.78	1.00			
WMBS	-0.02	0.03	-1.05	4.62	WMBS	0.64	-0.74	-0.88	0.73	1.00		
WMBG	-0.01	0.01	-0.43	3.72	WMBG	-0.10	0.13	0.11	-0.11	-0.17	1.00	
PMU	0.02	0.05	-0.57	5.50	PMU	-0.55	0.75	0.80	-0.87	-0.84	0.26	1.00

5 RESULTS

5.1 Sorts

Table 4 reports the monthly value-weight and equal-weight returns on 3 separate size portfolios, created with environmental, social, and governance score breakpoints. Size breakpoints are the 20th and the 50th market capitalization (share price times the number of common shares outstanding) percentiles for the micro caps, small caps, and big caps. E, S, and G breakpoints are the 30th and the 70th percentiles for the low, medium and high groups, determined with all-but micro firms. Initially, 9 portfolios for each E, S, and G factor are created. However, Table 4 also reports monthly returns for small, low, and high groups for the sample and the all-but micro size groups. Portfolios are formed at the end of each year and held until the end of next year. Accordingly, the table reports the monthly returns in excess of matching portfolios formed on size and book to market equity over the 2011-2019 period.

Results show that average monthly value weight returns are superior for low sorted E, S, G portfolios formed with the sample as well as all-but micro firms compared with the high sorted portfolios of the matching size. For example, the difference between the value-weighted returns on the high environmental score portfolios and low environmental score portfolios formed on all-but micro firms reach -1.28 percent per month, significantly different from zero at the 5 percent level with a t-statistic of -2.44. The same difference is -1.39 percent per month for portfolios formed on the social score and -1.82 percent per month for portfolios formed on governance score, both significant at the 5 percent level (t-statistic of -2.39 and -2.18 respectively). Put differently, a portfolio created by buying the low governance score firms and shorting the high governance score firms with an exposure to international markets would achieve on average a 22 percent per annum return over the 2011-2019 period.

Micro caps account for 23 percent of all sample firms (see Table 1) and a mere 0.94 percent of the sample market cap in this research. As a result, one can notice the impact of micro firms on sort portfolios. For example, monthly returns on equal weight high minus low portfolios formed with sample firms are not significantly different from zero, unlike the monthly excess returns on value weight sorts of the same size. For all – but micro portfolios one can notice that equal weight is still significant at the 5 percent level for the portfolios sorted on environmental and governance scores while the monthly excess returns for portfolios sorted on the social score are significant at the 10 percent level. All the high minus low monthly excess returns are higher for equal weight portfolios when compared with their value weight peers.

Table 4
Average Excess Returns for Portfolios Formed Using Variables of Interest

The table below shows the monthly average equal-weighted and value weighted returns from sort portfolios in excess of matching portfolios created on size and book-to-market equity over the 2011-2019 period. T-statistics show whether the means are significantly different from 0. Profitability is gross profits (Revenues-COGS) divided by total assets. Size breakpoints are the 20th and the 50th market capitalization percentiles. Sort breakpoints are the 30th and the 70th percentiles for the underlying variables, determined with all but micro stocks. Financials (GICS sector 40) are excluded.

	Low	**Medium**	**High**	**High - Low**	**Low**	**Medium**	**High**	**High - Low**
Panel A. Sorting on Environmental Score								
	Average Value – Weighted Returns				*t–statistics for Average Value Weighted Returns*			
Sample	0.75	0.07	-0.79	-1.68	1.76	0.35	-2.83	-2.19
Micro	0.47	0.12	0.13	-0.52	0.95	0.90	0.82	-0.63
Small	0.01	0.26	0.18	-0.33	0.01	1.96	2.42	-1.38
Big	-0.02	-0.21	-0.77	-0.91	-0.04	-1.59	-2.19	-1.80
All but micro	0.46	0.20	-0.76	-1.28	2.03	3.89	-2.81	-2.44
	Average Equal – Weighted Return				*t–statistics for Average Equal Weighted Returns*			
Sample	0.08	-0.04	-0.37	-0.45	0.15	-0.61	-0.79	-0.46
Micro	-0.72	-0.51	-0.50	1.60	-1.39	-4.52	-4.23	1.58
Small	0.12	0.11	-0.23	0.10	0.77	1.15	-1.96	0.47
Big	0.99	-0.41	-1.36	-0.81	1.54	-2.42	-3.03	-1.48
All but micro	0.34	0.24	-0.70	-1.09	1.96	3.79	-2.17	-2.12
Panel B. Sorting on Social Score								
	Average Value – Weighted Returns				*t–statistics for Average Value Weighted Returns*			
Sample	0.82	0.01	-0.79	-1.78	1.76	0.02	-2.93	-2.22
Micro	0.56	0.17	0.35	-0.64	0.99	1.35	1.16	-0.68
Small	-0.01	0.25	0.06	-0.45	-0.01	1.95	0.92	-2.05
Big	-0.04	-0.27	-0.75	-1.14	-0.09	-1.83	-2.22	-1.71
All but micro	0.56	0.11	-0.74	-1.39	1.83	1.92	-2.83	-2.29
	Average Equal – Weighted Return				*t–statistics for Average Equal Weighted Returns*			
Sample	0.16	-0.17	-0.33	-0.50	0.31	-1.22	-0.75	-0.51
Micro	-0.60	-0.38	-0.35	1.39	-1.11	-1.97	-1.25	-0.51
Small	0.04	-0.07	-0.30	0.07	0.14	-0.73	-1.79	0.21
Big	0.88	-0.30	-1.33	-0.93	1.55	-1.29	-3.00	-1.32
All but micro	0.37	0.18	-0.64	-1.06	1.44	1.78	-1.97	-1.75

Table 2 Continued

	Low	Medium	High	High - Low	Low	Medium	High	High - Low
Panel C. Sorting on Governance Score								
	Average Value – Weighted Returns				*t–statistics for Average Value Weighted Returns*			
Sample	0.78	0.16	-0.93	-1.88	2.02	1.68	-2.65	-2.32
Micro	0.19	0.29	0.69	-0.63	0.73	1.63	1.75	-1.27
Small	0.66	0.19	0.08	-0.72	2.30	2.72	1.19	-1.70
Big	-0.42	-0.40	-1.02	-1.87	-1.17	-1.68	-2.14	-1.96
All-but micro	0.74	0.10	-0.92	-1.82	1.99	2.20	-2.38	-2.18
	Average Equal – Weighted Return				*t–statistics for Average Equal Weighted Returns*			
Sample	0.10	-0.01	-0.42	-0.53	0.23	-0.07	-0.82	-0.56
Micro	-0.87	-0.03	0.10	1.63	-2.26	-0.11	0.26	2.00
Small	0.39	-0.03	-0.27	-0.67	0.72	-0.33	-2.48	-1.22
Big	0.91	-0.70	-1.50	-1.62	1.41	-2.22	-2.76	-1.66
All-but micro	0.58	0.20	-0.85	-1.53	1.75	2.32	-2.09	-1.95

5.2 Fama-Macbeth Regressions

Factor correlation matrix table (Table 3) showed the correlation coefficients across created RMRF, SMB, HML, WMBE, WMBS, WMBG, and PMU factors. Excluding WMBG, the correlation between the created factors was undeniably high. Accordingly, prior to the Fama-Macbeth, regressions variance inflation factors (VIF) for each discussed model are calculated to assess whether multicollinearity is an issue amongst the factor variables. These models (1) CAPM, (2) FF 3 factor model, (3) FF 3 Factor Model + PMU, are all tested for multicollinearity, including either an E S or G factors. Table A4 shows the VIF values for all 3 models including a WMBE, WMBS or WMBEG factor on all size groups. Recall that a VIF value of 10 and higher was an indicator of multicollinearity. It can be seen that the highest VIF value for any factor variable, which is a VIF value of 9.85, is observed within an FF 3 factor model with an added WMBE factor when employed with the micro caps. Yet, all the VIF values for all size groups are below 10. Thus, I decide to proceed with the FM regressions and run a separate FM regression for each model with an added E, S and G factor. Furthermore, I run a Breusch-Godfrey test and find that autocorrelation possesses an issue up to the second lag. Fittingly, I use 2 month lagged Newey-West standard errors in all the FM regressions to correct for the autocorrelation as well as possible heteroskedasticity of the residuals.

Table 5 displays the FM regressions run with Fama and French 3 factor model and an added WMBE, WMBS or WMBG factor for all size groups. Columns numbered 1,2 and 3 show the monthly average slopes from FM regressions run with an added WMBE, WMBS and WMBG factor respectively.

The intercept is significantly different from zero only in the small cap group when the FF 3 factor model includes an added environmental factor. In an FM model, an intercept significantly different from zero shows the monthly returns unexplained by the factor model employed. Considering the model performance for the micro and the small caps, the only variable that has significant explanatory power over monthly returns is the market factor. Yet, for the small caps, all the factors in FF 3 + WMBG is insignificant, highlighting the weak performance of the model.

Moreover, the redundancy of the value factor was shown in recent research, for example in Fama and French (2015). This research is somewhat in line with these findings. Overall, HML is significant in 3 out of 9 models employed, that is when combined together with the governance factor for the sample, big and all-but micro size groups. Amongst the E, S and G factors WMBS has significant explanatory power at 5 percent level when employed for the big and all-but micro firms. It is also significant at 1 percent level when run on the sample firms.

WBME is significant when run for the whole sample while the governance factor is significant for the all-but micro size group. All the E, S, G factors, when significant, have a positive sign in predicting monthly returns. Double sorts showed the difference between high-low E, S, G score portfolios whereas the WMBE, WBMS, WBMG factors were created with the difference between low and high E, S, G score portfolios. Accordingly, results in Table 5 can be seen as a robustness check for the double sorts.

Table 5
Fama-Macbeth Regressions

This table shows average slopes and their t-statistics from Fama-Macbeth (1973) regressions to predict stock returns. RMRF, SMB and HML are the Market, Size and Value factor mimicking portfolios. WMBE, WMBS, and WMBG are the worst minus best E, S and G factor mimicking portfolios. All factor portfolios except for RMRF are rebalanced in December of each year. RMRF is the monthly value weighted average excess return. Time period is 2011-2019. Micro, small and big size breakpoints are the 20th and the 50th market capitalization percentiles. Financials (GICS sector code 40) are excluded. T-statistics are in parentheses and corrected with Newey West standard errors of two lags.

	Sample			**Micro**			**Small**			**Big**			**All-but Micro**		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
RMRF	1.165*** (8.06)	1.226*** (8.76)	1.215*** (7.77)	1.789*** (2.66)	2.224*** (4.46)	1.743* (1.75)	1.335*** (2.98)	1.654** (2.38)	1.203 (1.63)	1.112*** (5.76)	1.065*** (5.41)	1.220*** (8.39)	1.030*** (4.64)	1.042*** (4.74)	1.202** (8.86)
SMB	-0.435* (-1.90)	-0.269 (-1.31)	-0.835*** (-5.80)	-0.986 (-1.02)	-0.266 (-0.36)	-0.449 (-0.52)	-4.029 (-1.57)	-0.106 (-0.06)	0.119 (0.12)	-0.575*** (-3.33)	-0.529*** (-3.87)	-0.726*** (-6.05)	-0.492*** (-2.82)	-0.447*** (-3.27)	-0.687*** (-5.65)
HML	-0.245 (-1.04)	-0.062 (-0.19)	-0.531*** (-2.78)	0.225 (0.19)	0.635 (0.82)	0.061 (0.04)	-1.844 (-1.28)	-1.258 (-1.05)	-1.464 (-1.09)	-0.232 (-0.79)	-0.35 (-1.06)	-0.495*** (-3.01)	-0.400 (-1.16)	-0.403 (-1.06)	-0.565*** (-3.59)
WMBE	0.598** (2.45)			-1.156 (-1.56)			-1.118 (-0.56)			0.405* (1.70)			0.424 (1.49)		
WMBS		0.811*** (4.05)			-0.441 (-0.74)			-0.038 (-0.02)			0.432** (2.15)			0.517** (2.19)	
WMBG			0.286 (0.81)			0.191 (0.29)			0.223 (0.41)			0.442* (1.72)			0.450* (1.83)
Int	0.005 (0.65)	0.008 (1.03)	0.009 (0.81)	-0.031 (-1.17)	-0.048 (-1.50)	-0.007 (-0.21)	0.112* (1.94)	0.065 (0.93)	0.030 (0.70)	-0.001 (-0.15)	0.003 (0.34)	0.006 (0.72)	-0.000 (-0.04)	0.006 (0.71)	0.007 (0.68)
Number of Obs	207,963	207,963	207,963	36,884	36,884	36,884	57,410	57,410	57,410	113,624	113,624	113,624	186,698	186,698	171,039
Adj R2	0.063	0.066	0.062	0.056	0.070	0.063	0.090	0.091	0.089	0.069	0.074	0.074	0.076	0.076	0.075

5.3 Regions in Focus: The US and Europe

Furthermore, also as a robustness check, I run FM regressions, employing all three models with an added WMBE, WMBS, and WBMG factor for the US and the Developed Europe regions. I specifically choose these regions since they account for more than 70% of sustainable investment assets in the world (Global Sustainable Investment Alliance). For my sample, these regions represent 59.2 percent of the total sample observations and 79.59 percent of the total market cap (Table A1). Moreover, Developed Europe's share across sample micro caps (35.58%), small caps (45.83) and big caps (28.15) is somewhat balanced while the US region dominates sample big caps (51.44) and accounts only for a small portion of sample micro caps (9.22%). Prior to the FM regressions, I run autocorrelation tests on all the models. I first run pooled OLS regressions containing all the independent variables, clustered on the stock level, and then robust Breusch-Godfrey tests. I find that autocorrelation does not possess an issue for the US sample. For the DE region results show standard errors need to be corrected up to the third lag. Accordingly, I corrected the Developed Europe Fama-Macbeth regression standard errors with Newey West 3 lags.

Average monthly slopes from the FM regressions for the US sample can be found in Panel A of Table 6. Columns 1 to 3 report the CAPM factor slopes, 4 to 6 the FF 3 factor model slopes, and 7 to 9 the FF 3 factor model plus the profitability factor slopes, all with an added WMBE, WMBS, and WMBG factor. When regressed separately with a CAPM model all E, S, and G factors have significant explanatory power over the average monthly returns. For CAPM with WMBG (column 3), the table reports an intercept (-0.028) negative in sign and significant at the 1 percent level (t-statistics of - 4.04). Additionally, column 3 is the only column where the governance score has a significant slope (-0.486). Unlike in the double sorts and the FM regressions of Table 5, the risk premium for the governance score in the US sample indicates a negative relationship between firm governance scores and firm excess returns. However, a significant intercept in column 3 indicates that this model is doing a poor job explaining the variation in the data. Indeed, when combined with more complex models (columns 6 and 9), this relationship fades, and the governance factor loses its significance.

For the US sample, the value premium is significant twice (columns 6 and 9) and carries a negative sign in both instances. A value slope of -1.028 in column 6 indicates that a one standard deviation increase in the value premium is expected to result in a 1.028 percent decrease in monthly excess returns when controlled for market and size and governance factors.

It can be seen in columns 4 and 5, neither size nor value has explanatory power while market, WMBE and WMBS slopes are highly significant and large in economic magnitude. While these findings are in line with those for the full sample, their economical implication appears to be higher for the US sample. Indeed, a one standard deviation increase in the US environmental and social score factors is anticipated to result in a 0.96 and 0.99 percent increase in monthly returns respectively. Moreover, factor premiums on the environmental score and governance score factors are both larger and have more explanatory power in the US sample when compared with the total sample. With a significant profitability factor, too WMBE and WMBS slopes are significant and positive in sign.

Moreover, columns 7 and 8 of Table 6, contradict the results of Novy-Marx (2013). While profitability is highly significant in explaining monthly excess returns, it is negative in sign, indicating a negative relationship between US firm profitability and monthly. In this research too, profitability has higher explanatory power than value.

Panel B of Table 6 reports the FM regressions on the Developed Europe sample. This section shows a completely different picture for E, S, and G as well as profitability when compared with the full sample and the US sample. To begin with, the slope on WBMG is not different from zero in all three models. Moreover, the factor premiums on E and S factors are highly significant for the CAPM model. Besides, in an FF 3 factor + WBME model, the E factor has a superior explanatory power on monthly returns when compared with the remaining regressors. For the social factor, when combined with the FF 3 factor model, a loss in explanatory power is observed. Yet in this model, WMBS remains significantly different than zero. These findings indicate that sample Developed European stocks with higher environmental and social scores have higher returns than stocks with lower environmental and social stocks. This contradicts the first and the second hypotheses that predict a higher return for lower scored E and S than firms with higher E and S scores respectively.

Duuren, Plantiga, and Scholtens (2016) indicated that the US portfolio managers were skeptical about ESG investing while the European portfolio managers were optimistic. Indeed, the results in this section might be seen as supplementary evidence on why the views on ESG differ across the US and the European portfolio managers.

Table 6
Fama and Macbeth Regressions for the US and Developed Europe

This table shows average slopes and their t-statistics from Fama-Macbeth (1973) regressions to predict stock returns. RMRF, SMB and HML are the Market, Size and Value factor mimicking portfolios. WMBE, WMBS, and WMBG are the worst minus best E, S and G factor mimicking portfolios. PMU is the profitability factor where profitability is gross profits (Revenues – COGS) divided by total assets. All factor portfolios except for RMRF are rebalanced in December of each year. RMRF is the monthly value weighted average excess return. Time period is 2011-2019. US is all firms in S&P 500. Developed Europe is Stoxx 600 firms except for those domiciled in Poland and Czechia. Financials (GICS sector code 40) are excluded. T-statistics are in parentheses and corrected with Newey West standard errors of three lags for the Developed Europe region.

Panel A: the US

	Dependent Variable Monthly Excess Returns ($r_i - r_f$)								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
RMRF	1.504*** (3.92)	1.647*** (2.68)	1.590*** (14.03)	0.898*** (4.33)	0.806*** (3.45)	0.926*** (8.18)	1.133*** (8.72)	1.201*** (7.34)	1.142*** (9.53)
SMB				-0.090 (-0.34)	-0.091 (-0.36)	-0.651*** (-5.78)	0.329** (2.26)	0.344** (2.12)	0.009 (0.03)
HML				-0.357 (-1.21)	-0.413 (-1.04)	-1.028*** (-5.95)	0.009 (0.05)	0.414 (1.10)	-0.690* (-1.80)
WMBE	1.742*** (3.92)			0.991*** (5.30)			0.565** (2.19)		
WMBS		2.004*** (3.00)			0.958*** (4.26)			0.885*** (3.19)	
WMBG			-0.486* (-1.83)			-0.283 (-1.12)			-0.386 (-0.60)
PMU							-0.735*** (-3.88)	-0.649*** (-3.37)	-0.637 (-1.38)
Int	0.023 (0.75)	0.047 (0.88)	-0.028*** (-4.04)	-0.013 (-1.17)	-0.014 (-1.20)	-0.011 (-1.51)	-0.008* (-1.75)	-0.006 (-1.33)	-0.005 (-0.81)
Number of Obs	51,453	51,453	51,453	51,453	51,453	51,453	51,453	51,453	51,453
Adj R2	0.190	0.190	0.169	0.192	0.192	0.189	0.197	0.197	0.197

Table 6 Continued

Panel B: Developed Europe

	Dependent Variable Monthly Excess Returns ($r_i - r_f$)								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
RMRF	0.284* (1.81)	0.410*** (2.60)	-0.311 (-0.59)	0.163** (2.20)	0.188 (1.62)	0.313** (2.06)	0.062* (1.65)	0.062* (1.65)	0.062* (1.75)
SMB				0.255*** (2.99)	0.295*** (3.04)	0.282*** (2.97)	0.089** (2.04)	0.084 (1.41)	0.089 (1.12)
HML				0.073 (0.91)	0.164* (1.94)	0.087 (0.54)	0.026 (0.86)	0.047 (1.30)	0.026 (1.52)
WMBE	-0.533*** (-3.16)			-0.306*** (-3.00)			-0.016 (-0.22)		
WMBS		-0.482** (-2.14)			-0.174* (-1.82)			-0.060 (-0.70)	
WMBG			-0.213 (-0.52)			-0.005 (-0.08)			-0.014 (-0.63)
PMU							0.416*** (4.04)	0.347*** (3.86)	0.416*** (3.03)
Int	0.015* (1.67)	0.007 (0.92)	0.025 (1.41)	0.005 (0.93)	-0.002 (-0.30)	0.008 (1.16)	0.000 (-0.16)	0.001 (0.18)	0.000 (0.00)
Number of Obs	72,794	72,794	72,794	72,794	72,794	72,794	72,794	72,794	72,794
Adj R2	0.000	0.000	0.000	0.011	0.011	0.011	0.008	0.008	0.008

When combined with the profitability factor, however, both WBME and WMBS lose their significance. In fact, profitability appears to have more explanatory power on monthly returns than any other factor variable. In columns 7,8 and 9 profitability factor is significant at the 1 percent level and has a positive sign. The positive sign shows higher profitability explains higher returns for the DE sample firms, as opposed to the US and full sample firms. Put differently, a one standard deviation increase in the profitability factor is associated with a 0.41 percent increase in monthly returns.

5.4 Environmental, Social and Governance Scores on Valuation Metrics

Table 7 displays the results from robust month fixed effects for valuation metrics, Tobin's q (columns 1 to 3), Price-to-Earnings ratio (columns 4 to 6), and market to book ratio (columns 7 to 9) using size, industry beta, environmental, social and governance scores as independent variables.

For the US sub-sample, the E and G scores appear to have predictive power on future Tobin's q. Yet this finding does not carry a significant economic magnitude. Indeed, while significant at the 5 percent level, both E and G score slopes in columns 1 and 3 indicate that a unit increase in these scores is expected to decrease Tobin's q by 0.002 and 0.001 percentage points respectively. On the other hand, the relationship between social score and Tobin's q appears to be insignificant for the US firms. For the European sub-sample, all three E, S, and G scores appear to carry significant explanatory power over the future value of the firm. Like the US sub-sample, higher E and G scores predict a negative lower Tobin's q. Unlike the US sub-sample, the social score has a highly significant explanatory power over Tobin's q in the European sample.

The relationship between ESG and price-to-earnings ratio seems to be missing for both the US and the European sample. Instead, the relationship between E, S and G scores and future market to book ratio appears to be significantly and positively correlated. This relationship is stronger for the European sub-sample, both in significance and in economic magnitude. Specifically, a unit increase in social score is associated with an 0.005 percent higher book to market equity for the European sub-sample. On the other hand, a unit increase in social score translates with an 0.003 percent higher book to market equity for the US sample.

It is important to highlight that the market-to-market ratio in Table 7 can be seen as a more simplified version of Tobin's q. Accordingly, due to the opposing results in the slopes of these valuation metrics and the insignificant relationship between P/E ratio and E, S and G scores, I am not able to display an economically significant evidence for the relationship between firm fundamentals and E, S, and G scores.

Table 7
Average Slopes on Valuation Metrics

Table 7 shows average slopes from the monthly fixed effects regressions on valuation metrics of the US and the European sub-samples. The dependent variables are the natural logarithm of Tobin's q, the natural logarithm of Price to Earnings (P/E) ratio, and the natural logarithm of the ratio of market to book equity (MB). LnMC is the natural logarithm of market capitalization (price times shares outstanding). Beta is the 12-month rolling window industry beta, calculated from value weighted sub-sample market returns. LnMC, E, S and G scores are lagged 12 months. The test period is 2011-2019. Financials (GICS sector code 40) and observations with negative P/E and MB ratios are excluded. Robust t-statistics are in parentheses.

Panel A: US (S&P 500)

	Dependent Variable lnTobin's Q			*Dependent Variable lnP/E*			*Dependent Variable lnMB*		
	(1)	(2)	(3)	(4)	(5)	(6)	*(7)*	*(8)*	*(9)*
lnMC (t-12)	0.300***	0.296***	0.697***	0.327***	0.323***	0.320***	-0.452***	-0.441***	-0.431***
	(15.21)	(15.49)	(15.15)	(10.42)	(10.31)	(10.68)	(-11.84)	(-11.89)	(-11.77)
Beta	-0.005	-0.008	-0.009	-0.156***	-0.156***	-0.156***	0.010	0.009	0.014
	(-0.23)	(-0.36)	(-0.40)	(-3.66)	(-3.66)	(-3.66)	(0.21)	(0.19)	(0.30)
Env Score (t-12)	-0.002**			-0.001			0.004***		
	(-2.31)			(-0.56)			(2.87)		
Social Score (t-12)		-0.001			0.000			0.003*	
		(-1.62)			(-0.16)			(1.79)	
Gov Score (t-12)			-0.001**			0.000			0.002*
			(-2.54)			(0.40)			(1.79)
Int	-2.273***	-2.242***	-4.464***	-0.004	0.013	0.010	2.924***	2.855***	2.805***
	(-12.37)	(-12.15)	(-11.86)	(-0.02)	(0.05)	(0.04)	(8.47)	(8.36)	(8.12)
Number of Obs	44,055	44,055	44,055	43,316	43,316	43,316	46,095	46,095	46,095
R2	0.250	0.248	0.249	0.065	0.064	0.064	0.172	0.168	0.168

Table 7 continued

Panel B: Europe (Stoxx 600)									
	Dependent Variable lnTobin's Q			*Dependent Variable lnP/E*			*Dependent Variable lnMB*		
	(1)	(2)	(3)	(4)	(5)	(6)	*(7)*	*(8)*	*(9)*
lnMC (t-12)	0.275***	0.279***	0.272***	0.312***	0.306***	0.319***	-0.303***	-0.320***	-0.295***
	(12.83)	(12.91)	(12.59)	(10.42)	(10.31)	(10.68)	(-11.84)	(-11.89)	(-11.77)
Beta	-0.001	-0.001	-0.001	-0.095**	-0.095**	-0.094**	0.001	0.001	0.001
	(-0.85)	(-0.86)	(-0.85)	(-3.66)	(-3.66)	(-3.66)	(0.21)	(0.19)	(0.30)
Env Score (t-12)	-0.002*			0.001			0.004***		
	(-1.98)			(-0.56)			(2.87)		
Social Score (t-12)		-0.002***			0.001			0.005***	
		(-2.72)			(0.74)			(5.30)	
Gov Score (t-12)			-0.001**			-0.001			0.003***
			(-2.25)			(-0.39)			(4.08)
Int	-2.106***	-2.13***	-2.110***	0.114	0.144	0.125	1.887***	1.961***	1.855***
	(-11.33)	(-11.40)	(-11.35)	(-0.02)	(0.05)	(0.04)	(8.47)	(8.36)	(8.12)
Number of Obs	58,115	58,115	58,115	52,167	52,167	52,167	59,483	59,483	59,483
R2	0.137	0.138	0.137	0.030	0.030	0.030	0.079	0.083	0.077

6 CONCLUSION

In this paper, using an international dataset that covers the 2010-2019 period, I investigated whether a higher environmental, social or governance score translated into higher stock market returns and whether ESG affected firm valuation.

Primarily, I evaluated the variation in average monthly returns from E, S, and G score portfolios created following the double sorts approach similar to that in Fama and French (2008). Largely, my results showed higher monthly returns for lower E, S, and G scored value weighted and equal weighted portfolios. The difference in average monthly returns between all-but micro E, S and G hedge portfolios, that is low-high portfolios, amounted to 1.28, 1.39 and 1.89 percent respectively. Furthermore, these findings are significant at the 5% level, and in excess of matching portfolios created on size and book to market equity.

Then, I evaluated whether E, S and G scores explained average stock returns with Fama-Macbeth (1973) regressions. My initial findings were to some extent in line with those in the double sorts approach. I found a negative predictive power of E and S scores over monthly returns for my full sample, and negative predictive power of S and G scores over monthly returns for my all-but micro sample.

These results enabled me to answer the first three hypotheses of my research. My first two hypotheses which stated that 1) a lower environmental score and 2) a lower social score predicts higher average monthly returns for the full sample could not be rejected. Conversely, the third hypothesis that a lower governance score predicts higher average monthly returns for the full sample was rejected.

Furthermore, I employed Fama-Macbeth (1973) regressions, for the US and European sub samples, and found contradicting results between these two sub-samples. Similar to the full sample, the US firms showed a positive risk premium for E and S scores, even when controlling for Fama and French 3 factors plus the profitability factor of Novy-Marx (2013). These findings are in line with Pedersen, Fitzgibbons, and Pomorski (2020) which also shows a negative relationship between E as well as S scores and monthly returns. However, while Pedersen, Fitzgibbons, and Pomorski (2020) find a positive relationship between the governance score and the monthly returns, the same relationship in this research is not significant.

On the contrary, for the European firms, I documented a negative relationship between E as well as S scores and average monthly returns. This can rather explain why US portfolio managers are skeptical and their European peers are optimistic about ESG's implications on financial performance as highlighted by Duuren, Plantiga, and Scholtens (2016). Moreover, for the European sub-sample the E and S scores became redundant in explaining the monthly returns once controlled for the profitability factor of Novy-Marx (2013). My fourth hypothesis stated that environmental, social and governance scores predict the stock market returns negatively in the US sub-sample and positively in the European sub-sample. While a contradiction was indeed found on the environmental and social scores, for the sub-samples the governance score was not significantly different than zero. Therefore, the fourth hypothesis was rejected.

Lastly, for the US and European sub-samples, I investigated whether the findings in stock returns translated into firm fundamentals. Using Tobin's q as a proxy for firm value and controlling for size and industry beta, I discovered that firms with higher E and G scores were relatively undervalued in the US sub-sample when compared with firms with higher E and G scores, while the relationship between the social score and Tobin's q was insignificant.

For the European sample, the same pattern was observed for the E and G scores. Additionally, the results indicated that a higher social score amongst European firms predicted a future lower valuation. It is important to note that these results are negligible in economic magnitude and contradict the findings of Li et. al. (2018) which predicts a positive relationship between ESG and Tobin's q, and Surroca et. al. (2020) which documents a positive relationship between firm social performance and Tobin's q. Furthermore, I reject my fifth hypothesis which stated a positive relationship between E, S and G scores and Tobin's q.

Last of all, I evaluated whether E, S, and G scores carried significant explanatory power over price to earnings and market to book ratios. Accordingly, I found an insignificant relationship between all three E, S and G scores, and a significantly positive relationship between E, S, and G stocks and market to book equity. My results differ from Hong and Kacperczyk (2009) that display lower MB and PE ratios for sin stocks, and Pedersen, Fitzgibbons, and Pomorski (2020) as well as Cao et. al. (2019) that find an underpricing for low ESG stocks.

6.1 Limitations and Future Research

The biggest limitation in this research was due to the COVID-19 pandemic. The pandemic reduced my access to various databases substantially. This, in the end, extended my data gathering and cleaning processes and caused me to conduct my research with fewer than anticipated observations.

Unlike financial data, ESG data is not audited and disclosed on a voluntary basis. At the same time, the quantifiable part of ESG data is fairly limited. Accordingly, ESG ratings in databases are determined in conjunction with the disclosed data and surveys conducted by these rating agencies (Refinitiv, 2021). However, as mentioned in the introduction, the proportion of sustainable investing to conventional investing and the availabilityof the ESG data continues to increase, and this increase is expected to continue (PWC, 2020). Accordingly, this will likely increase the statistical power of future research on ESG and firm financial performance.

Regarding my analyses, it is important to note that when compared with the monthly price data, the number of observations in firm fundamentals is significantly smaller since most fundamentals are reported annually. Accordingly, with data covering a limited time period, the strength of my analyses on firm fundamentals is relatively weak. Certainly, conducting these tests with a more extensive database remains an interesting area for future research.

Although of the regions in my research was Asia, the ESG data was not sufficient and I could not employ a separate test for the Asian sub-sample. However, it is known that the share of sustaianble invesment in Asia is growing rapidly (Global Sustainable Investment Alliance, 2018). Therefore, this region will be certainy noteworthy for future research.

7 Bibliography

Aouadi, A., & Marsat, S. (2018). Do ESG Controversies Matter for Firm Value? Evidence from International Data. *Journal of Business Ethics, 150*, 1027-1047.

Baker, M., Bergstresser, D., Serafeim, G., & Wurgler, J. (2018). Financing The Response To Climate Change: The Pricing and Ownership of U.S. Green Bonds. *Nationl Bureau of Academic Research Working Paper*.

Banz, R. (1981). The Relationship Between Return and Market Value of Common Stocks. *Journal of Financial Economics, 9*(1), 3-18.

Barber, B., Morse, A., & Yasuda, A. (2021). Impact Investing. *Journal of Financial Economics, 139*(1), 162-185.

Benabou, R., & Tirole, J. (2011). Identity, Morals, and Taboos: Beliefs as Assets. *The Quarterly Journal of Economics, 126*(2), 805-855.

Bergstresser, B., & Philippon, T. (2006). CEO Incentives and Earnings Management. *Journal of Financial Economics, 80*(3), 511-529.

Bernstein, A., Gustafson, T. M., & Lewis, R. (2019). Disaster on the Horizon: The Price Effect of Sea Level Rise. *The Journal of Financial Economics, 134*(2), 253-272.

Cao, J., Titman, S., Zhan, X., & Zhang, W. (2019). ESG Preference and Market Efficiency: Evidence from Mispricing and Institutional Trading. Working Paper.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance, 52*(1), 57-82.

Chung, H. K., & Pruitt, S. P. (1994). A Simple Approximation of Tobin's q. *Financial Management, 23*(3), 70-74.

De Bondt, W. F., & Thaler, R. (1985). Does the Stock Market Overreact? *The Journal of Finance, 40*(3), 793-805.

Delis, M. D., de Greiff, K., Iosifidi, M., & Ongena, S. (2019). Being Stranded with Fossil Fuel Reserves? Climate Policy Risk and the Pricing of Bank Loans. *Swiss Finance Institute Research Paper No. 18-10*.

Duuren, E., Plantiga, A., & Scholtens, B. (2016). ESG Integration and the Investment Management Process: Fundamental Investing Reinvented. *The Journal of Business Ethics, 138*(3), 525-533.

Engle, R., Giglio, S., Kelly, B., Lee, H., & Stroebel, J. (2020). Hedging Climate Change News. *The Review of Financial Studies, 33*(3), 1184-1216.

Fama, E., & MacBeth, J. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy, 81*, 607-636.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25*(2), 383-417.

Fama, E. F., & French, K. R. (1993). Common Risk Factors in the Returns of Stocks and Bonds. *Journal of Financail Economics, 33*, 3-56.

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 427-465.

Fama, E. F., & French, K. R. (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics, 116*(1), 1-22.

Giese, G., Lee, L., Melas, D., Nagy, Z., & Nishikawa, L. (2019). Foundations of ESG Investing: How ESG Affects Equity Valuation, Risk, and Performance. *The Journal of Portfolio Management, 45*(5), 69-83.

Global Sustainable Investment Alliance. (2018). *Global Sustainable Investment Review.* Global Sustainable Investment Alliance.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica, 46*(6), 1251-1271.

Hong, H., & Kacperczyk, M. (2009). The Price of Sin: The Effects of Social Norms on Markets. *Journal of Financial Economics, 93*(1), 15-36.

IMF: World Economic Outlook (April, 2019) - GDP, Current Prices. (2019). From IMF: https://www.imf.org/external/datamapper/NGDPD@WEO

Jegadeesh, N., & Titman, S. (1993). Returns on Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance, 48*(1), 65-91.

Kishan, S., & Massa, A. (2021, March 18). *Bloomberg.* From https://www.bloomberg.com/news/articles/2021-03-18/blackrock-to-press-companies-on-human-rights-and-nature?srnd=green-finance

Krüger, P. (2015). Corporate Goodness and Shareholder Wealth. *The Journal of Financial Economics, 115*(2), 304–329.

Li, Y., Gong, M., Zhang, X., & Koh, L. (2018). The Impact of Environmental, Social, and Governance Disclosure on Firm Value: The role of CEO power. *The British Accounting Review, 50*(1), 60-75.

Lindenberg, E. B., & Ross, S. A. (1981). Tobin's q Ratio and Industrial Organization. *Journal of Business, 54*(1), 1-32.

Lintner, J. (1965). Security Prices, Risk, and Maximal Gains from Diversification. *The Journal of Finance, 20*(4), 587-615.

Mooney, A. (2020, November 16). *Financial Times.* From https://www.ft.com/content/dd01aacd-85a0-4577-9700-26f1d6fb26b3#comments-anchor

Morck, R., Shleifer, A., & Vishny, R. (1988). Management Ownership and Market Valuation: An empirical analysis. *The Journal of Financial Economics, 20*, 293-315.

Novy-Marx, R. (2013). The Other Side of Value: The Gross Profitability Premium. *Journal of Financial Economics, 108*(1), 1-28.

PWC. (2020). *2022 - The growth opportunity of the century* . From https://www.pwc.lu/en/sustainable-finance/esg-report-the-growth-opportunity-of-the-century.html

Pastor, L., Stambaugh, R., & Taylor, L. (2019). Sustainable Investing in Equilibrium. Working Paper.

Pedersen, L. H., Fitzgibbons, S., & Pomorski, L. (2020). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, Working Paper.

Ponkratz, N., Bauer, R., & Dervall, J. (2019). Climate Change, Firm Performance, and Investor Surprises. Working Paper.

Principles for Responsible Investment. (2020). *PRI Update Q4.*

Refinitiv. (2021). *Refinitiv*. From https://solutions.refinitiv.com/esg-data

Rosenberg, B., Kenneth, R., & Lanstein, R. (1985). Persuasive Evidence of Market Inefficiency. *Journal of Portfolio Management, 11*(3), 9-16.

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

Shiller, R. J. (2003). From Efficient Markets Theory to Behavioral Finance. *Journal of Economic Perspectives, 17*(1), 83-104.

Surroca, A., Aguilera, R., Desender, K., & Tribo, A. J. (2020). Is Managerial Entrenchment Always Bad and Corporate Social Responsibility Always Good? A cross-national examination of their combined influence on shareholder value. *Strategic Management Journal, 41*(5), 891-920.

Tobin, J. (1969). A General Equilibrium Approach to Monetary Theory. *The Journal of Money, Credit and Banking, 1*(1), 15-29.

United Nations. (2015, December). Paris Agreemet. 1-27. Paris: United Nations.

Wernerfelt, B., & Montgomery, C. (1988). Tobin's q and the Importance of Focus in Firm Performance. *The American Economic Review, 78*(1), 246-250.

World Bank. (2021). From https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?most_recent_value_desc=true&locations=US

World Meteorological Organization. (2020, December). From https://public.wmo.int/en/media/news/2020-closes-decade-of-exceptional-heat

APPENDIX

Table A1

Percentage Number of Observations and Market Capitalization Across Regions

Table A1 shows region and country share in the number sample observations, in all size groups and for the total sample size where size equals market capitalization (MC). Sample period is 2010-2019.

	Percent Sample number of Obs.	Percent Sample Micro Cap	Percent Sample Small Cap	Percent Sample Big Cap	Percent Sample total MC
Sample Asia-Pacific	**32.37**	**48.41**	**25.92**	**16.20**	**16.96**
Australia	*17.17*	*33.44*	*7.85*	*2.23*	*2.85*
New-Zealand	*2.12*	*4.30*	*1.83*	*0.01*	*0.17*
Hong Kong	*2.91*	*0.34*	*2.27*	*5.46*	*5.14*
Japan	*10.17*	*10.33*	*13.97*	*8.50*	*8.80*
USA	**23.32**	**9.22**	**20.12**	**51.44**	**48.22**
Sample DE	**35.88**	**35.58**	**45.83**	**28.15**	**29.11**
Continental DE	*26.26*	*18.32*	*32.84*	*22.59*	*23.00*
UK	*9.62*	*17.26*	*12.98*	*5.56*	*6.11*
Sample EM	**8.43**	**6.80**	**8.13**	**4.21**	**5.71**
EM AP	*0.15*	*0.04*	-	*1.26*	*1.15*
Emerging Europe	*0.35*	*0.06*	*0.40*	*0.14*	*0.15*
South Africa	*1.57*	-	-	-	*1.27*
Brazil	*4.39*	*5.74*	*5.49*	*1.75*	*2.01*
Mexico	*1.97*	*0.95*	*2.25*	*1.07*	*1.13*

Table A2

the US and Europe Means by Sector

MC is market capitalization. Profitability is gross profits (revenues – COGS) divided by total assets. BME is ratio of book equity to market equity. Tobin's Q is measured for the US and European samples only, Sample period is 2010 – 2019.

	Number of Firms	**% of Sub Sample MC**	**E Score**	**S Score**	**G Score**	**Profita -bility**	**BME**	**Tobin's Q**
Energy	105	11.72	53.27	56.76	58.36	0.39	1.19	1.67
Materials	160	6.36	61.57	61.43	57.66	0.66	1.54	1.32
Industrials	222	12.13	51.26	56.98	54.17	0.66	0.60	1.59
Consumer Discretionary	169	10.85	48.83	57.42	54.41	0.67	0.55	1.91
Consumer Staples	112	14.21	60.11	63.90	56.14	0.83	0.61	2.17
Healthcare	140	15.95	44.18	59.67	56.60	0.41	0.53	2.69
Information Technology	141	13.10	43.44	56.15	53.90	0.44	0.44	2.68
Communication Services	90	9.80	46.72	56.81	51.81	0.63	0.74	1.37
Utilities	48	3.61	61.92	59.16	66.86	0.77	0.78	0.78
Real Estate	62	2.27	41.88	52.92	49.28	0.09	1.91	6.22

Table A3
Correlation Coefficients Across Regions

Profitability is gross profits (revenues – COGS) divided by total assets. BME is ratio of book equity to market equity. Sample period is 2010 – 2019.

Panel A: Asia – Pacific

	Market Cap	Environ-mental Score	Social Score	Governance Score	Profitability	BME
Market Cap	1.000					
Environmental Score	0.287	1.000				
Social Score	0.278	0.672	1.000			
Governance Score	0.180	0.440	0.489	1.000		
Profitability	-0.011	-0.012	-0.017	-0.010	1.000	
BME	0.048	0.469	0.174	0.072	-0.012	1.000

Panel B: The - US

	Market Cap	Environ-mental Score	Social Score	Governance Score	Profitability	BME
Market Cap	1.000					
Environmental Score	0.303	1.000				
Social Score	0.321	0.730	1.000			
Governance Score	0.154	0.426	0.352	1.000		
Profitability	0.037	-0.039	0.067	-0.044	1.000	
BME	-0.125	-0.009	-0.087	0.022	-0.282	1.000

Panel C: Europe

	Market Cap	Environ-mental Score	Social Score	Governance Score	Profitability	BME
Market Cap	1.000					
Environmental Score	0.331	1.000				
Social Score	0.329	0.716	1.000			
Governance Score	0.253	0.327	0.385	1.000		
Profitability	-0.097	0.001	-0.018	0.040	1.000	
BME	-0.080	0.053	0.018	-0.020	0.003	1.000

Panel D: Emerging Markets

	Market Cap	Environ-mental Score	Social Score	Governance Score	Profitability	BME
Market Cap	1.000					
Environmental Score	0.138	1.000				
Social Score	0.180	0.759	1.000			
Governance Score	0.102	0.423	0.471	1.000		
Profitability	-0.037	-0.001	0.035	0.077	1.000	
BME	0.142	0.145	0.125	0.038	-0.096	1.000

Table A4
Variance Inflation Factors

Model (1) is CAPM, (2) is Fama and French 3 factor model, (3) Fama and French 3 factor model + PMU (profitable minus unprofitable factor) of Novy-Marx (2013), (4) is model (3) excluding value factor. E, S, and G refer to WBME, WBMS, WMBG, worst minus best factors created for the underling E, S and G score pillars. Details about the models can be found in the methodology section.

Panel A: All Regions

	(1) + E	(1) + S	(1) + G	(2) + E	(2) + S	(2) + G	(3) + E	(3) + S	(3) + G
RMRF	1.45	1.70	1.01	1.54	1.72	1.53	1.53	1.75	1.53
SMB				2.39	2.39	2.13	2.59	2.62	2.54
HML				6.80	4.60	2.56	6.80	4.83	3.62
WBME	1.45	1.70	1.01	7.25			9.71		
WBMS					5.55			6.51	
WMBG						1.02			1.11
PMU			4.78				4.52	3.96	3.66

Panel B: Micro Cap

	(1) + E	(1) + S	(1) + G	(2) + E	(2) + S	(2) + G	(3) + E	(3) + S	(3) + G
RMRF	1.46	1.71	1.01	1.55	1.73	1.55	1.55	1.77	1.55
SMB				2.46	2.45	2.15	2.65	2.68	2.59
HML				6.84	4.64	2.51	6.84	4.88	3.72
WBME	1.46	1.71	1.01	7.35			9.85		
WBMS					5.60			6.64	
WMBG						1.02			1.11
PMU							4.69	4.11	3.77

Panel C: Small Cap

	(1) + E	(1) + S	(1) + G	(2) + E	(2) + S	(2) + G	(3) + E	(3) + S	(3) + G
RMRF	1.46	1.72	1.01	1.55	1.74	1.55	1.55	1.77	1.55
SMB				2.41	2.39	2.12	2.62	2.65	2.58
HML				6.71	4.57	2.56	6.71	4.78	3.59
WBME	1.46	1.72	1.01	7.21			9.67		
WBMS					5.54			6.54	
WMBG						1.02			1.11
PMU							4.59	4.04	3.73

Panel D: Big Cap

	(1) + E	(1) + S	(1) + G	(2) + E	(2) + S	(2) + G	(3) + E	(3) + S	(3) + G
RMRF	1.43	1.69	1.01	1.51	1.71	1.51	1.51	1.74	1.51
SMB				2.37	2.37	2.12	2.56	2.59	2.51
HML				6.93	4.59	2.54	6.83	4.83	3.61
WBME	1.43	1.69	1.01	7.22			9.67		
WBMS					5.53			6.45	
WMBG						1.02			1.10
PMU							4.45	3.88	3.60

Panel E: All-but Micro

	(1) + E	(1) + S	(1) + G	(2) + E	(2) + S	(2) + G	(3) + E	(3) + S	(3) + G
RMRF	1.44	1.70	1.01	1.52	1.72	1.52	1.52	1.75	1.52
SMB				2.38	2.38	2.12	2.58	2.61	2.53
HML				6.79	4.59	2.55	6.80	4.82	3.60
WBME	1.44	1.70	1.01	7.23			9.67		
WBMS					5.54			6.48	
WMBG						1.02			1.11
PMU							4.49	3.93	3.6

Table A5
Yearly Breakpoints

MC is market capitalization in millions of US dollars. BME is book to market equity. E, S and G are environmental, social and governance scores. Pr is profitability where profitability is gross profits (Revenues – COGS) divided by total assets.

YEAR	MC20	MC50	BM30	BM70	E30	E70	S30	S70	G30	G70	Pr30	Pr70
2010	1803	4916	0.34	0.84	33.02	68.23	36.54	65.39	39.88	68.74	0.31	0.78
2011	1527	4452	0.38	1.29	32.45	68.32	37.14	66.53	39.40	69.67	0.30	0.75
2012	1577	4943	0.36	1.23	33.97	69.29	38.26	67.22	40.30	69.22	0.30	0.74
2013	1734	5717	0.31	1.11	33.20	68.99	37.80	66.88	40.64	68.83	0.29	0.73
2014	1656	5917	0.30	1.12	34.18	68.40	39.09	67.45	40.70	68.67	0.28	0.70
2015	1512	6149	0.29	1.13	35.82	69.93	42.30	69.67	43.01	69.70	0.28	0.71
2016	1583	6349	0.32	1.23	38.83	70.92	45.21	71.39	44.59	70.38	0.26	0.68
2017	1933	7687	0.30	1.12	39.71	71.80	47.59	73.76	45.33	70.61	0.25	0.65
2018	1697	6929	0.34	1.34	42.21	73.83	50.40	75.80	47.55	72.25	0.25	0.64
2019	1878	7711	0.31	1.25	46.71	75.89	54.55	78.20	49.35	73.58	0.22	0.60