ESG: The Relationship Between “Ethical” Investing and Abnormal Returns

Author: Per Malm

Master Essay II – 15 ECTS

Supervisor: Hossein Asgharian

Department of Economics - Spring 2022

Abstract

This essay examines the relationship between ESG and abnormal returns and its implications on investing. To investigate this topic, I allocate stocks into zero-investment portfolios based on high and low ESG, using three different weighting methods, equal weighting, value weighting and portfolio optimization. The portfolios are then evaluated using the CAPM as well as the Fama-French three-factor and five-factor models. In addition, I also calculate the Sharpe ratios and mean ESG scores of each individual sub portfolio, as well as the mean excess return of each zero-investment portfolio. The sample consists of 510 US stocks and covers the period between August 2009 and November 2019. As a measurement for ESG, I use a score for each pillar and an overall ESG score.

The results in general show that there is little to no impact of ESG on abnormal returns. There does seem to be some impact when we use value weighted portfolios. This might be due to a few outliers that either under- or overperform, as the abnormal returns persist even when controlling for the size effect. In practical terms this means that investors can generally expect adequate risk-adjusted returns even when preferring high ESG stocks. However, because of the indication that there may be outliers this should be taken with some degree of caution. Regarding the Fama and French factors there appears to be a presence of a value effect related to ESG. Broadly, low ESG stocks tend to be riskier as the high minus low portfolios indicate that they are more exposed to the factor loadings.

Keywords: ESG, Abnormal Returns, High Minus Low, Asset Pricing Model

"…say my name…"

Table of contents

1. Introduction

The general purpose of this essay is to investigate the effect that ESG has on stock returns and to understand the potential implications that this has for an investor seeking to form portfolios based on ESG. More specifically to see if there are any discrepancies between high and low ESG stocks in terms of abnormal returns. To determine whether there is a relationship between ESG and financial performance. This topic can be viewed both in a broader economic perspective in terms of its implications on "ethical investing", but also useful for investors seeking risk arbitrage opportunities by trading on ESG. As investors in the latter category can implement a similar investment strategy. Further, the purpose of this paper is also to see if the investment results improve by using portfolio optimization. Which will also indicate if there is a difference between the returns of optimal portfolios of high and low ESG, in terms of abnormal returns. As this has not, to the best of my knowledge, been examined in earlier literature.

ESG, environmental social corporate governance, has seen substantial growth as a field in terms of both customer and investor preferences. And thereby also the amount of capital invested in ESG assets. This has not gone unnoticed by firms as the number of companies reporting ESG data increased from around 20 companies in the early 1990s to about 9.000 in 2016. On the investor side networks and organizations have sprung into existence, such as the Principles for Responsible Investment (PRI), a UN initiative, which aims to promote ESG amongst institutional investors (Amel-Zadeh & Serafim, 2018). As of 2021 PRI has attained signatures from 3826 institutional investors with a total of 121,3 trillion dollars under management (UNPRI, 2022).

Studies regarding the economics of preferences that override pure profit motives go back to works such as Becker (1957). Becker modeled and studied the effects of discrimination in the economy wherein agents, such as employers, choose to adhere to social norms rather than purely maximize profits. In Becker's model agents take on financial costs because of limiting their selection of potential employees. This can be likened to current day investments in ESG wherein investors limit, or entirely overlook, investments in certain industries for example. Hong and Kacperczyk (2009) investigate the effects of what they call "the price of sin", namely the effect of investing in, or overlooking, certain industries based on social norms.

Hong and Kacperczyk classify certain stocks as "sin stocks" based on their industry classification, such as gambling, alcohol, and tobacco. They find that sin stocks generate higher expected returns and abnormal returns compared to non-sin stocks. And, that investors constricted by norms, such as pension funds, are less inclined to invest in sin stocks compared to investors such as hedge funds.

More recently, Pedersen et al. (2021) take a broader approach and investigate ESG by, amongst other topics, sorting portfolios based on ESG metrics. More specifically they create zero-investment portfolios wherein they take a long position in high ESG stocks and a short position in low ESG stocks. The authors use a mixture of ESG proxy variables, such as greenhouse gases for the environmental factor, but also an overall ESG score. To evaluate the portfolios, the authors look at the abnormal returns, generated using a few different asset pricing models. In conclusion they find that the portfolios tend to generate significant abnormal returns, indicating that there is a discrepancy between high and low ESG stocks. The results of Halbritter and Dorfleitner (2015) indicate that ESG scores have a negligible, to no impact on abnormal returns. Statman and Glushkov (2009) are also rather inconclusive in terms of the potential of investment strategies social responsibility ratings to generate abnormal returns. Earlier studies in contrast, such as Kempf and Osthoff (2007) find some positive relationship between some social responsibility ratings and abnormal returns. When sorting on overall ESG scores in specific, both Pedersen et al. (2021) as well as Halbritter and Dorfleitner (2015) find no significant abnormal returns.

To investigate this topic, I form zero-investment portfolios based ESG scores of each pillar as well as an overall ESG score. Namely, one score for E, S, G and ESG, thereby considering a broader perspective of ESG rather than only looking at specific proxy variables. As these ratings integrate multiple different aspects of a corporation's impact on the environment, etc. The basic idea of a zero-investment, or high minus low, portfolio is that the overall weights cancel each other out, and so, an equally large position is taken in both the long and short sub portfolios. Thereby indicating if the returns have a bias based on the sorting (Alexander, 2000).

To test the returns generated by the portfolios I apply various asset pricing models, namely the capital asset pricing model (CAPM), the Fama-French three-factor model (FF3), and the Fama-French five-factor model (FF5). This will indicate whether the sorting causes any substantial abnormal returns and is a popular method when testing the performance of portfolios in general.

Further, the methodology of creating zero-investment portfolios and evaluating them using Fama-French type models is particularly popular when evaluating the impact of ESG on stock returns (Pedersen et al., 2021). The portfolio formation and testing are executed in MATLAB, through an unreasonable number of loops.

Given previous literature the purpose and methodology of this essay are in the spirit of Pedersen et al. (2021) as well as Halbritter and Dorfleitner (2015) amongst others. There are three main contributions to the existing literature. First, this paper will in some sense fill a gap by looking at a more recent period, in this case 2009-2019. Further, a standard approach to weighting the sub portfolios is using either equal-weighted or value-weighted stocks or both. Whereas I also use portfolio optimization in addition to the standard method, which to the best of my knowledge has not been tried in previous literature. Finally, and arguably most importantly, the empirical results of previous literature in terms of what implications ESG has on stock returns is rather mixed. Pedersen et al. (2021) for example show that the sorting on G generates positive significant abnormal returns, whereas Bebchuk et al. (2013) find that this effect eventually disappears. Halbritter and Dorfleitner (2015) tend to find insignificant results for all ESG factors. This contrasts with Pedersen et al. (2021) who find it for not only G but also and to some extent E and S, depending on method. With different authors generating different results. I thereby hope that this paper will provide further insight into the topic by weighing in additional research.

The main results of my paper show that there is little to no impact of ESG on abnormal returns. Indicating that there is an absence of discrepancies between the returns of high and low ESG stocks. This result suggests that investors with ESG preferences can expect adequate risk-adjusted returns, because of the absence of mispricing. Although this generally holds true the results of the portfolios based on value weights suggest that there may be multiple outliers related to ESG. Stocks that either under- or overperform who can have a significant impact when portfolios are not diversified. Further, my essay also proposes that because of the non-significant abnormal alphas arbitrageurs will not be able to profit on ESG by implementing a similar investment strategy. They should instead focus on implementing strategies specifically targeting outlier firms if they are indeed present. Finally, it seems that using optimal weighted sub portfolios does not improve the performance of the high minus low portfolios. And that there are no discrepancies in returns between optimal portfolios of relatively high and low ESG.

The essay has the following outline: In chapter 2 I investigate previous research related to ESG, both from a theoretical and empirical perspective, particularly articles who also employee asset pricing models. Chapter 3 describes the method in detail and motivates the choices behind it by discussing and referring to previous research on related topics. Chapter 4 accounts for the data that I use, the choices behind it, and some basic descriptive statistics. In Chapter 5 I present and discuss the results in terms of abnormal returns, factor loadings, sub portfolio Sharpe ratios and mean ESG scores as well as the mean excess returns of each zero-investment portfolio. Finally, in chapter 6 I make my concluding remarks based on the results and the analysis of them. Chapter 7 includes the reference list.

2. Previous Research

This section is an overview of the previous literature on the topic of ESG investing. First, the theoretical background gives insight into what “ought” to happen in a world of ESG investing and the implications of various asset pricing models. Secondly, the empirical literature review is divided into sections based on every ESG aspect. As investigations of the effects of ESG on stock returns do not necessarily have to be constrained to using zero-investment portfolios and/or ESG scores, the empirical literature review covers both papers focused on high minus low type ESG strategies and papers using other approaches. The final section describes the CAPM, FF3 and FF5.

2.1 Theoretical Background: Equilibrium Asset Pricing Models and the ESG-Efficient Frontier

According to CAPM, all investors are fully informed and hold the market portfolio, meaning that they hold all available assets in their portfolio each of which is value weighted. Further, this means that there are no abnormal returns as all assets will be on the security market line (SML) (Sharpe, 1964). Later equilibrium asset pricing models has sought to relax the assumptions of the CAPM, one famous example of this is arbitrage pricing theory, introduced by Ross (1976). Merton (1987) introduced an equilibrium model which takes into consideration incomplete information. In this model Merton relaxes the assumption that all investors hold all available assets, instead investors are assumed to operate under incomplete information. Which in this case means that investors simple do not know about all available assets. Meaning that they completely exclude some subset of the available assets in their optimal portfolio. As pointed out by Merton, his model can in principle be extended to other scenarios in which, for whatever reason, investors choose to exclude some set of assets. And so, this should give us some insight into what might happen in a scenario where some investors exclude assets because of ESG metrics.

In the incomplete information model assets that are “neglected” will have a higher equilibrium expected return relative to if there was complete information. This is because these assets will not receive adequate investments, a case of underinvestment.

This of course also means that they will be undervalued relative to their hypothetical equilibrium value under complete information (if investors invested in all assets). Further, this means that the SML falls apart as some assets will have a $\alpha_j \neq 0$. This in turn also means that the market portfolio will no longer be efficient (Merton, 1987).

Heinkel et al. (2001) introduced an equilibrium model wherein some investors exclude firms they believe to be unethical, in this case firms that are polluting. In this model green investors choose to exclude all polluting firms from their optimal portfolio, whereas the neutral investors include all available assets. The effect that the existence of green investors has on the equilibrium pricing in this setting depends primarily on three different factors, the number of green investors, the cost for a polluting firm to reform into a non-polluting one, and the correlation between polluting and non-polluting cash flows [1].

First, if the number of green investors grows, the amount invested in polluting firms decreases, leading to a lower stock price. This also increases the expected return in equilibrium and thereby the cost of capital. The increased cost of capital creates an incentive for polluting firms to reform into non-polluting firms. Whether or not a firm will follow through with this reform also depends on the cost of reforming, if too high relative to the cost of not reforming, they will not and vice versa. Further, the cost of reforming will be reflected in the stock price for a reformed firm. If a firm reforms, then its new price will simply be its price while it was a polluter plus the cost of reforming. Meaning that the green investors must be willing to pay the higher stock price for the firm to want to reform. Finally, the correlation between polluting and non-polluting cash flows influences the equilibrium pricing by affecting how green and neutral investors optimize their portfolios. If the correlation is close to one, in absolute terms, there are greater possibilities for diversification meaning that neutral investors can hold more of polluting and non-polluting firms. Which in turns creates less incentives for reform as well as lower equilibrium expected returns for polluting firms, as there will be less underinvestment in polluting firms (Heinkel et al., 2001).

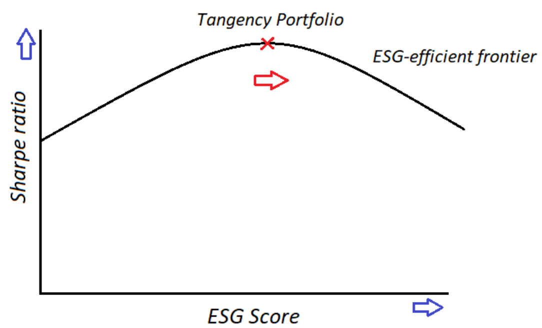
Zerbib (2020) builds an asset pricing model that follows the framework originally developed by Merton (1987).

[1] There are further technical details that determine equilibrium prices in this model; however, I believe that describing these are of no value here. For a full derivation of the model see the original article by Heinkel, Kraus and Zechner (2001).

In this model investors are either regular; meaning that they invest in all available assets, sustainable; meaning that they exclude certain assets based on ESG, or integrators; who invest in all assets, but weight the assets based on ESG. As in the model by Heinkel et al. (2001) the least ethical firms and or environmentally friendly end up with a higher cost of capital and equilibrium expected return.

Luo and Balvers (2017) model the empirical observation that sin stocks tend to have abnormal returns and show that such an effect can indeed be modelled. The authors argue that this effect occurs because neutral investors, that want to hold the market portfolio, need to hold a disproportionate amount of the sin stocks. This entails a higher risk through concentration in sin stocks, which they in turn want compensation for.

The theoretical and empirical results of ESG investing is mixed, with some authors arguing that taking ESG into account should necessarily result in lower returns, whereas some argue that ESG based investing outperforms other strategies. Pedersen et al. (2021) develop a theory that they argue can reconcile these different perspectives on ESG investing. The traditional way or reasoning around investments in financial theory is the mean-variance framework, originally developed my Markowitz (1952). Wherein assets should be thought of in terms of benefits, expected returns, and the downside, volatility. This led to the development of performance measures such as the famous Sharpe ratio (Sharpe, 1966). However, in a world where investors may also consider ESG when investing, this framework arguably must be expanded upon. Pedersen et al. (2021) show that this relationship, between returns, volatility and ESG can be broken down into a much simpler relationship. Namely, a relationship between the Sharpe ratio and ESG of a portfolio. In this model an investor no longer only puts returns in relation to risk, as in the traditional framework. But rather Sharpe ratio in relation ESG in what the Pedersen et al. (2021) calls “the ESG-efficient frontier”. This can be shown in the following simple illustration:



The tangency portfolio is the same as in the standard mean-variance framework, it does not take ESG scores into account but is simply the portfolio with the highest Sharpe ratio. Investors that are indifferent to ESG and instead simply seek the optimal portfolio can pick the tangency portfolio as per usual. Those investors who consider both the Sharpe ratio and ESG will stay to the right of the tangency portfolio, with those having stronger preferences for ESG staying further to the right. The right-side is the ESG-efficient frontier because investors on the left of the overall frontier can improve either their Sharpe ratio or ESG score. Therefore, there is no reason for any investors to stay to the left of the tangency portfolio (Pedersen et al., 2021).

Pedersen et al. (2021) further expand upon this theory by considering an ESG-adjusted CAPM to predict equilibrium prices and returns in a world with both regular and ESG investors. This model predicts a few different outcomes depending on investor behavior. Perhaps somewhat unsurprisingly, when there are many investors that simply seek to maximize their Sharpe ratio and do not care for ESG, the prices of ESG stocks simply generate expected returns in line with their ability to generate profits and so on. Meaning that ESG has no effect. In the opposite scenario, when there are many investors that care about ESG, the expected returns of high ESG stocks drop as these investors are willing to give up some of their profit in favor of ESG. Further implying that the cost of capital for firms with a high ESG also decreases. This has some interesting implications as a lower cost of capital would mean that these firms now have a greater amount of net positive investments. Which could in turn increase the profitability of the firms and drive up the expected returns. This could explain, on a theoretical level, why some empirical results suggest that high ESG stocks outperform low ESG stocks. The opposite would of course happen for firms with a low ESG in this scenario. With many investors preferring high ESG these firms would instead see their expected returns increase and thereby also their cost of capital. In summary, it all depends on investor preferences.

2.2 Empirical Background: The Environmental Factor

Pedersen et al. (2021) investigate, amongst other topics, the abnormal returns and the average excess returns of zero-investment portfolios sorted on greenhouse gas emissions. Where emissions is considered as a proxy for the environmental factor E. Stocks from the S&P 500 index are sorted based on their emissions wherein in a long position is taken in high ESG stocks, meaning those with low emissions, and a short position is taken in low ESG stocks.

Further, the stocks are either equally weighted or value weighted. First, the average excess returns are positive for both the value weighted and equally weighted versions of the portfolio. However, only the average excess return of the value weighted portfolio is found to be significant. Next, the authors investigate the abnormal returns as defined by α_j using four different asset pricing models, namely, the CAPM, FF3, FF5 and FF6 (FF + momentum). Their portfolios yield positive abnormal returns when using all pricing models. However, there is only spurious significance amongst the portfolios. The FF5 and FF6 generate a significant α_j for both the equally and value weighted portfolios, FF3 generates no significant α_j for any portfolio and CAPM only generates a significant α_j for the equally weighted portfolio. These results are not entirely consistent, they do however suggest that portfolios sorted on E generate positive abnormal returns.

Halbritter and Dorfleitner (2015), like Pedersen et al. (2021), use zero-investment portfolios sorted on ESG to investigate its effect on returns. In this case the authors use ESG scores rather than proxy variables to conduct the sorting and the Carhart (1997) four-factor model to test for abnormal returns. The results of Halbritter and Dorfleitner (2015) when it comes to E, are rather inconclusive. There is no consistency in terms of whether the abnormal returns generated by the portfolios are positive or negative, or even significant. It also varies depending on the subsample, whether the portfolios are equal-weighted or value-weighted and which ESG score that they employ. In conclusion they find no relationship between E and abnormal returns.

Derwall et al. (2005) create zero-investment portfolios based on social responsibility ratings related to E. Using the CAPM they find that the portfolio generates positive and relatively large abnormal returns, however not significant. When testing the portfolio using a multifactor model, they also find a relatively large, but significant abnormal returns. Kempf and Osthoff (2007) use the Carhart four-factor model to test equal-weighted portfolios sorted on social responsibility ratings. They find that their environmentally friendly sorted portfolios generate positive and significant abnormal returns whereas their long-short portfolios have positive, but not significant abnormal returns.

Haan et al. (2012) sort portfolios based on corporate environmental performance (CEP). They find a negative relationship between CEP and stock returns, and further that low CEP portfolios tend to generate higher abnormal returns when applying multifactor models.

Dutta et al. (2018) investigates the relationship between returns and volatility and CO2 emissions as well as clean energy stocks. The authors find a low correlation between CO2 emissions and stock prices. Eccles et al. (2014) show that both high and low sustainability firms generate positive and significant abnormal returns, with the high sustainability firms generating the highest. While using the Fama-French four-factor model plus momentum. Bolton and Kacperczyk (2021), find a positive effect of CO2 on risk-adjusted returns. According to Amel-Zadeh and Serafeim (2018) more and more companies report emissions, suggesting that the environment is becoming an increasing concern for investors.

In summary, as the previous literature described above indicates, there is no clear relationship between stock returns and abnormal returns and the environmental impact of a firm. Most studies do however find that high-E stocks tend to generate positive abnormal returns, although this is not consistently significant. It is of course important to keep in mind that these studies vary in their methodological approach, their choice of environmental variable and data. It is also noteworthy that these studies focus on the US.

2.3 Empirical Background: The Social Factor

As briefly discussed in the introduction, Hong and Kacperczyk (2009) investigate the relationship between returns and what they define as sin stocks, namely those firms involved in alcohol, tobacco, and gambling. They thereby, by proxy, investigate the social aspect of ESG by defining the social factor (S) as what is perceived to be socially acceptable. Because there is a social norm against alcohol, tobacco, and gambling, firms that operate in these industries are basically seen as having a low S. This can also be observed in who holds which stocks in practice. For example, pension funds, whom generally need to adhere to certain social norms tend to hold less sin stocks. Whereas hedge funds for example, who are arbitrageurs, do not follow any social norms and thereby hold relatively large positions in sin stocks. To classify their stocks, they use industry indices and create portfolios over a relative long timespan, 1926-2006. They find that both during the 1926-2006 period as well as the 1965-2006 period sin-stocks have positive and significant abnormal returns. Both when applying a single factor CAPM-style model and when using Fama-French type multifactor models, using additional factors such as small minus big (SMB) and momentum etc.

Further, sin stocks are found to have higher expected returns compared to otherwise equivalent non-sin stocks. Luo and Balvers (2017) apply a similar approach and also find a sin stock premium. In contradiction to these two papers Blitz and Fabozzi (2017) find that once a more adequate asset pricing model is used, the FF5, this effect disappears.

Pedersen et al. (2021), when forming zero-investment portfolios sorted on S, apply the methodology of Hong and Kacperczyk (2009). Namely using an industry index to define stocks as having either a high or a low S. Their results are consistent with those of Hong and Kacperczyk for their value-weighted portfolio. This portfolio generates negative abnormal returns when applying all models, it is however not significant when applying the FF5 and FF6. Further, the average excess returns are negative and significant for this portfolio, again indicating that sin stocks outperform non-sin stocks. For their equal-weighted portfolio their results do however paint a different picture. The equal-weighted high S minus low S portfolio generates positive, but not significant, average excess returns. When it comes to the abnormal returns of this portfolio the results vary depending on which model is applied. The CAPM α is negative, but not significant, a result that is relatively consistent with the value-weighted portfolio. However, for the FF3, FF5 and FF6 the abnormal returns are positive, but not significant. What this suggests is that there may be some sin stocks with a relatively high value that have a major impact on driving the portfolio returns and thereby also the results. Therefore, when the stocks are equally weighted, this effect most likely disappears as the weights are no longer dependent on the market cap.

Halbritter and Dorfleitner (2015), when sorting on S, use a score as when sorting on E. Their high minus low portfolios are not consistent when sorted on S either. The abnormal returns vary depending on sub portfolio, both when it comes to the ESG score provider and weighting scheme. Some of the portfolios have a positive and some have a negative α_j, and not all are significant. They conclude that there is no relationship between S and abnormal returns. Kempf and Osthoff (2007) in addition to sorting portfolios on an environmental factor, also sort them on a social factor, using social responsibility ratings. In the case of S, they use a few different ratings. They use ratings for community, human rights, diversity, and product. For the first three ratings the long-short portfolios generate positive abnormal returns, however only the community one is significant. The product long-short portfolio generates negative, but not significant, abnormal returns.

The fact that the product portfolio generates negative abnormal returns is consistent with Pedersen et al. (2021) and Hong and Kacperczyk (2009). Although Kempf and Osthoff (2007) use a rating, it is nonetheless in this case based on the product that the firm produces and so it is reasonable to suggest that it is closely related to which industry the firm operates in. And so, firms in alcohol, gambling and tobacco should receive relatively low ratings, comparable to being defined as sin stocks.

2.4 Empirical Background: The Corporate Governance Factor

Bebchuk et al. (2013) investigate the effect of governance on abnormal returns. They find that the once positive, in the 1990s, correlation between the two disappeared in the 2000s. Good governance with respect to ESG is not necessarily as affected by social norms, as with sin and non-sin stocks. Because an investor that disregards social norms may still desire to invest in firms with good governance as this should arguably positively impact returns. Bebchuk et al. argue that this is precisely why the abnormal returns disappeared, because investors learned that good governance is important for returns. This has been further investigated with respect to the cost of equity and debt by Zhu (2014) who finds that good governance is associated with lower cost of equity and debt. Meaning that investors are more willing to invest in corporations with good governance and that they believe that this is less risky, relative to companies with poor governance. Because of this corporations themselves have an incentive to implement good governance as it reduces the cost of capital and therefore leads to more net positive present value investment opportunities.

Going back once more to some of the articles previously described in the sections on the social and environmental pillars of ESG, many of these also investigate the governance aspect. First, Pedersen et al. (2021) use low accruals as a proxy variable, suggesting that firms with more conservative accounting practices are better governed (Sloan, 1996), (Kim et al.,2012). Out of all portfolios that Pedersen et al. (2021) form, the once formed on G are by far the most consistent. In this case the authors find that portfolios formed on low accruals generate positive and significant abnormal returns using all asset pricing models and both equal-weighted and value-weighted sub portfolios. The results of Halbritter and Dorfleitner (2015) when sorting on the governance factor are, as with the other factors, inconsistent when it comes to generating α_j.

Their result varies not only depending on the source of their ESG rating as well as their weighting method, but also within each of these. Kempf and Osthoff (2007) sort portfolios on an employee relations rating. Their results show that stocks with high employee relations ratings generate positive, and for the most part significant α_j.

2.5 The CAPM and the Fama & French Three/Five Factor Model

The Capital asset pricing model and the Fama-French three- and five-factor models can all be used to test assets for abnormal returns. However, the foundation as well as empirical support vary between them (Fama & French, 2015). The CAPM has its foundation in financial theory, more specifically, mean-variance theory. The CAPM assumes that all investors hold the market portfolio, which consists of all available assets, in equilibrium, and, that these investors are all mean-variance optimizers with the only thing differentiating them being their risk aversion. Further, all investors are assumed to be able to borrow or lend money at the same terms, what Sharpe (1964) calls “a common pure rate of interest”. Further, all investors have homogenous expectations about $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the expected returns and covariances. All of this means that, according to the CAPM, the only difference in investors’ portfolio composition is how much they choose to hold in the risk-free asset and the market portfolio. What this means in practical terms for an investor is that the only thing being priced is systematic risk, or how much market risk an investor is willing to take on. The market risk is thereby the only thing being priced, and so the expected return of an asset can be determined using the following equation (Sharpe, 1964):

$$E(R_j) = R_f + \beta_m[E(R_m) - R_f] \tag{1}$$

The FF3 on the other hand has its basis in empirical research, namely that of the inadequacy of the CAPM to sufficiently explain stock returns in practice. As can be noted in equation 1, no intercept term exists, this is because in accordance with theory there should be no α_j. This is because this term represents a mispricing and if the assumptions of the CAPM hold true, no mispricing should exist. However, empirical research has shown that the CAPM tends to precisely generate this form of mispricing. Meaning the market factor alone is not sufficient to explain asset pricing (Fama & French, 1993).

This led to the development of the FF3 by Fama and French (1993), who introduced two additional factors which they find remedy the problems of the CAPM. These factors are small minus big (SMB), which takes the size effect into account. The size effect means that firms with a low market capitalization tend to outperform firms with a high market capitalization. Something that Fama and French hypothesis could be a result of underinvestment, as investors might overlook small firms. The authors form this factor according to the following equation:

$$\begin{aligned} SMB = \frac{1}{3}(Small\,Value + Small\,Neutral + Small\,Growth) \\ - \frac{1}{3}(Big\,Value + Big\,Neutral + Big\,Growth) \end{aligned} \tag{2}$$

Where SMB is the average return of this zero-investment portfolio, where the long position is of course represented by three sub portfolios with small stocks and vice versa for big. The second additional factor is high minus low (HML) which considers the value effect. This effect is related to the fact that high-book-to-market value stocks tend to outperform low-book-to-market value stocks. Where the latter is seen as growth firms, firms that receive investments on the promise of future value. HML is formed as such:

$$HML = \frac{1}{2}(Small\,Value + Big\,Value) - \frac{1}{2}(Small\,Growth + Big\,Growth) \tag{3}$$

Where HML, like SMB, is the average return of a zero-investment portfolio (Fama & French, 1993). The FF3 has been shown to outperform the CAPM in explaining stock returns in several empirical papers, for multiple periods and countries. Amongst these are Fama and French (1996, 1998), Gaunt (2004) and Taneja (2010) for example. The regression equation of the FF3 can be written as:

$$R_{j,t} - R_{f,t} = \alpha_j + \beta_{m,j}\left[R_{m,t} - R_{f,t}\right] + \beta_{s,j}SMB_t + \beta_{h,j}HML_t + e_{j,t} \tag{4}$$

The FF5 was introduced by Fama and French in 2015 who, based on research by Titman et al. (2004) amongst others, identify two additional factors. Fama and French (2015) show that augmenting the original FF3 will lead to a model that more adequately explains stock returns. The new factors are robust minus weak (RMW), and conservative minus aggressive (CMA). RMW is the returns of a zero-investment portfolio of stocks with robust and weak profitability.

CMA is the difference in returns between conservative and aggressive firms, defined as firms with high and low level of investment respectively. Another thing of note is that Fama and French also slightly alter the SMB and HML factors, calling the second orthogonal HML (HMLO). The regression equation of the FF5:

$$R_{j,t} - R_{f,t} = \alpha_j + \beta_{m,j}\left[R_{m,t} - R_{f,t}\right] + \beta_{s,j}SMB_t + \beta_{h,j}HMLO_t + \beta_{r,j}RMW_t + \beta_{c,j}CMA_t + +e_{j,t} \tag{5}$$

3. Method and Research Framework

This chapter describes and motivates the method and models that I apply and some basic background of these in the context of previous research. I start by explaining how the portfolios are formed and how they are weighted. Next, I explain how the portfolio results are evaluated.

3.1 Forming and Weighting High minus Low ESG Portfolios

High minus low portfolios, or zero-investment portfolios, are formed by taking a long position in a group of stocks and an equally large short position in another. From an investment point of view these two portfolios essentially cancel each other out, meaning that a hypothetical investor needs no initial investment. Hence the name zero-investment, as the short position finances the long position. The basic idea of forming high minus low, or zero-investment, portfolios is, as briefly mentioned in the introduction, to see whether these portfolios generate any significant abnormal returns. If so, this will indicate a discrepancy in the returns between the two sub portfolios. Therefore, this will signal if there is a difference in returns between the two different group of stocks, for example, those stocks with high ESG scores and those with low ESG scores. This can be defined by the following simple equation as high minus low (HML) (Alexander, 2000):

$$r_{L,t} - r_{S,t} \equiv HML_t \tag{6}$$

Where, $r_{L,t}$ is the return of the long sub portfolio and $r_{S,t}$ is the return of the short sub portfolio at time *t*. This method has been employed extensively to test for various stock market effects. Debondt and Thaler (1985 & 1987) use it to investigate the so-called momentum effect in stock returns. Seyhun (1986) uses it to investigate insider trading and its implications for information-efficiency in the market. La Porta et al. (1997) investigates value stocks in the context of market-efficiency. As described in the chapter 2, high minus low portfolios have also been extensively used to test the effect of ESG on stock returns.

First, I use a cut-off point of 50 stocks in each of the sub portfolios, basically going long in 50 stocks and short in 50 stocks. My total sample consists of 510 stocks meaning that these sub portfolios consist of roughly the 10% highest and lowest ESG scored stocks respectively.

The choice of 50 stocks in specific is somewhat arbitrary and there are different possible cut-off points that could be considered. The advantage of using 50 stocks is that you potentially get relatively large difference between the ESG scores of the two portfolios, compared dividing them into bigger groups. Meaning that it is easier to capture mispricing. At the same time the number of assets is large enough for having diversified portfolios. Halbritter and Dorfleitner (2015) use multiple different cut-off points, ranging from 1% to 50%. Pedersen et al. (2021) use a single quantile (cut-off point) when sorting their portfolios, with a sample also consisting of roughly 500 stocks. Kempf and Osthoff (2007) use a cutoff of 33%. The main purpose of my essay is to test if ESG has an actual effect on stock returns and therefore I believe that it is redundant to use multiple cut-off points.

Next, to weight the sub portfolios I use three different methods, equal-weighted, value-weighted and mean-variance optimized weights. The first two are consistent with previous research when it comes to forming ESG high minus low portfolios. They will simply capture slightly different effects as the value-weighted sub portfolios consider market cap whereas equal-weighted sub portfolios disregard it. Pedersen et al., 2021 and Halbritter et al. (2015) use both equal-weighted and value-weighted sub portfolios whereas Kempf and Osthoff (2007) only use equal-weighted. I will give a brief overview of portfolio optimization in the next subsection.

All portfolios are reweighted on a monthly basis meaning that each month a new sorting is conducted based on that months ESG scores. The equal-weighted sub portfolios do of course remain the same for each period, each asset has a 2% weight. Whereas the value-weighted sub portfolios are reweighted each month as the market capitalization changes. The portfolios are held for the following month after each sorting, generating a monthly return as per equation 6, and then a new sorting is conducted and so on. This procedure is executed between August 2009 and November 2019, with the final return being generated on the final trading day of November 2019. Yielding a total of 124 monthly returns for the equal- and value-weighted high minus low portfolios. If a stock is delisted during the sample period, it is simply removed from the next month's sorting. For the optimized sub portfolios this procedure is a bit more intricate, this will be described in the next section.

3.2 Portfolio Optimization

Originally introduced by Markowitz (1952), the basic idea of portfolio optimization is to maximize expected returns in relation to volatility. The reasoning behind this has its origin in the mean-variance framework developed by Markowitz, wherein if you pick between two assets with the same standard deviation, the one with the highest expected return should always be preferred and vice versa. This basic concept led to the introduction of the Sharpe ratio as a portfolio performance measure by Sharpe (1966):

$$Sharpe\ ratio_j = \frac{E(R_j) - R_f}{\sigma_j} \tag{7}$$

Where a higher Sharpe ratio is preferred, as this gives a higher expected return, $E(R_j)$, in relation to risk (standard deviation), σ_j. Sharpe ratio is not only a popular performance measure, but it can also be used to define optimal portfolio weights. Portfolio optimization is conducted by picking the asset weights that maximize the Sharpe ratio. This is illustrated by the following optimization problem, along with the two constraints that I use:

$$\begin{gathered} \max_{\boldsymbol{w}}\ Sharpe\ ratio_j = \max_{\boldsymbol{w}} \frac{\boldsymbol{w}'\boldsymbol{\mu}}{\sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}^{-1}\boldsymbol{w}}} \\ s.t.\ \boldsymbol{w}'\mathbf{1} = 1 \\ s.t.\ 0 \leq w_i \leq 0.1 \end{gathered} \tag{8}$$

Where $\boldsymbol{\mu}$ is a $Nx1$ vector containing the expected excess returns of each asset, $\boldsymbol{w}$ is a $Nx1$ vector containing the weight of each asset, $\boldsymbol{\Sigma}$ is a NxN covariance matrix of the excess returns and finally w_i is the weight of asset i. The basic idea behind why this might be desirable is the same as that behind why Sharpe ratio is a valid performance measure. Pick the weights that maximize returns in relation to risk for the portfolio. Which in theory, under certain assumptions, should also be the portfolio which maximizes the utility for a risk averse investor (Danthine & Donaldson, 2014). An example of the implementation of portfolio optimization in conjunction with ESG is the ESG-efficient frontier created by Pedersen et al. (2021), which I briefly described in the previous chapter. In this case the authors maximize the Sharpe ratio with an additional ESG constraint, namely by targeting a specific ESG score while simultaneously maximizing the Sharpe ratio.

The optimized sub portfolios are weighted using the following methodology. As with the previous two weighting schemes I sort the assets each month in the sample period depending on their respective ESG scores. However, in this case I start at a later period, namely three years after the first observation, in august 2012. The reason for this is that I use a sample of 36 monthly excess returns to estimate $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ for each sub portfolio every month. Meaning that every month I use the excess returns of the previous 36 months to generate the optimal weights, with equation (8). This portfolio is then, like with the other two, held for one month before it is reweighted using the stocks of the new sorting. For example, in each time, t, I sort the 50 stocks with the highest and lowest ESG scores. I then use the excess returns of those stocks going back 36 months, to t-36, to estimate $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ which gives me the optimal weights for each sub portfolio individually, for the high and low stocks. At the end of the month these sub portfolios each generate a return which is weighted against each other in the zero-investment portfolio.

Another thing worth considering is the constraint that I place on the weights of the sub portfolios. First, and relatively straight forward is $\boldsymbol{w}'\mathbf{1} = 1$, which basically means that the individual asset weights should sum to 100%. This is basically saying that 100% of your money should be invested in the portfolio. Keeping the total weight constraint at 100% makes sure that the results are consistent as the equal-weighted and value-weighted sub portfolios do of course also have a total portfolio weight of 100%.

The second and arguably most important constraint, $0 \leq w_i \leq 0.1$, has two meanings. First, no individual stock should be shorted, each individual weight must be positive. Secondly, no stock should have a higher weight than 10%. The reasoning behind having a short constraint is related to the high minus low investment strategy itself. The whole idea is to investigate going long in high ESG scored stocks and short in low ESG scored stocks. For example, if an individual stock is shorted in the low ESG portfolio, this will mean that in the high minus low portfolio, a long position is taken in this stock. So not having a short sale constraint would defy the purpose of the paper. The 10% weight limit constraint is, to a certain extent, a judgement call as well as a practical consideration. One inherent weakness of the classical portfolio optimization approach is that it is very much input sensitive. Meaning small shifts in the estimated excess returns and covariance matrix can greatly affect the resulting weights. The estimated excess returns can have a particularly large effect. This can cause some assets to gain a disproportionate weight in the final portfolio. Meaning that a few assets can have a relatively large effect on the portfolio performance.

The misspecification of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in combination with disproportionate portfolio weights can lead to a sub optimal portfolio. Basically, meaning that there exists another combination of assets which leads to a higher Sharpe ratio. This creates the necessity of having a weight constraint when implementing the model in practice (Chopra & Ziemba, 1993) (Black & Litterman, 1990). One possible remedy to this problem is using Bayes-Stein shrinkage estimation for example, instead of just using the sample excess returns and covariance matrix (Jorion, 1986). However, this is beyond the scope of this paper. It is also questionable how relevant it is as the point of this paper is to compare high and low ESG stocks, and not absolute portfolio performance. And so, having a weight constraint is in and of itself well-grounded in previous research, whereas the choice of 10% is somewhat arbitrary (Best & Grauer, 1991). There are of course other reasonable options rather than 10%.

Normally only equal and value weighted stocks are tested and not optimal sub portfolios. So, why test it with optimized portfolios as well? A common method to forming portfolios is using some type of portfolio optimization in the process (Bodie, Kane & Marcus, 2014). And so, optimal sub portfolios are seemingly more realistic than having equal- and value-weighted sub portfolios. The abnormal returns of the high minus low portfolio will in this case basically indicate if there is a discrepancy in the returns of optimal portfolios based on high and low ESG. Further, because this sub portfolio should, at least in theory, have the highest possible expected return in relation to risk. This might improve the performance of the high minus low portfolios. As the CAPM, FF3 and FF5 of course take risk in account when measuring α_j. If there is discrepancy between high minus low stocks in the first place of course.

3.3 Portfolio Evaluation

In the context of testing ESG stocks for abnormal returns the CAPM, FF3 and FF5 are all very much relevant as can be noted in the chapter on previous research. Another possible option is the Cahart four-factor model (Cahart, 1997). For example, Halbritter and Dorfleitner (2015) as well as Kempf and Osthoff (2007) use the Cahart four-factor model whereas Pedersen et al. (2021) use the CAPM, FF3, FF5 and FF6. Further, time series data, cross sectional data or a combination of both are all common and viable options. With either regular OLS or Fama and Macbeth (1973) type regressions. In this paper I only consider time series data and apply OLS to estimate the parameters of the CAPM, FF3 and FF5.

With previous research in mind this seemed like an obvious choice as these models are common both in testing for mispricing in the context of ESG but also in general. In practice I straightforwardly regress the monthly returns of each of the high minus low portfolios using the factors of all three models respectively. This will indicate if the abnormal returns are negative or positive, a negative or positive intercept, or whether such a mispricing even exists in the first place, an insignificant intercept. Which would of course mean a bias, or lack of bias, toward high or low ESG stocks, showing whether there is a difference in returns based on ESG score (Pedersen et al., 2021). Thereby generating clear results with respect to the purpose of my paper.

In addition to the asset pricing models, I also calculate the Sharpe ratios of each sub portfolio, by applying equation (7). Further, I also calculate the overall average ESG scores of each sub portfolios using the following equation:

$$\bar{S} = \frac{1}{T}\sum_{t=1}^{t=T} \boldsymbol{w_t} \boldsymbol{S_t'} \tag{9}$$

Where $\boldsymbol{S_t}$ is a $1xN$ vector containing the sub portfolio ESG scores each t and $\boldsymbol{w_t}$ is a $1xN$ vector containing the sub portfolio weights each t. The equal-weighted sub portfolios will have ESG scores that are proportional to the ESG scores of the included stocks. Whereas the optimal and value-weighted sub portfolios may place different weights on high and or low ESG stocks. And so, the average ESG score will indicate whether the weighting of the sub portfolios will have a bias in and of itself, in addition to the sorting. For example, in the high sub portfolio, a stock with a relatively large market value but relatively low ESG score may shift average ESG score downwards.

4. Data and Descriptive Statistics

This chapter gives a brief overview and description of the data that I use as well as the reasoning behind the choices that I make. I also present some basic descriptive statistics of the ESG scores and discuss these in the context of the paper.

4.1 Sample and Data

The sample period is August 2009 until November 2019 in terms of ESG scores. The first observation is for the final trading day of July 2009 as I use the final trading price of July and August 2009 to calculate the monthly return of August 2009 and so on. The reason for choosing this specific period is threefold, first and most importantly availability of data, secondly, it is a relatively "calm" period, and thirdly it is a recent period. I used ESG scores from Sustainalytics who provide ESG scores for a relatively large amount US firms, both private and public. Sustainalytics coverage gradually improves over time. So, I had to choose between a larger number of firms in the sample and a longer sample period. I ended up choosing August 2009 as this gave me a final sample of 510 US companies. Further exclusion of stocks was based on the criteria that they should be public from the beginning of the period. As previously mentioned, stocks that are delisted after August 2009 are simultaneously removed from the sorting. Meaning that the sample technically becomes somewhat smaller throughout the period. A sample of this size is consistent with previous papers, like Pedersen et al. (2021) who look at S&P 500 stocks for example. With their ESG scores being updated monthly, a period of roughly ten years should give enough observations to conduct reasonable inference on the sample.

With respect to the second reason for choosing this period, 2009 to 2019 is relatively calm, or "consistent". As it begins right after the financial crisis of 2008 and ends right before COVID-19. And so, no extreme event should have a major impact on the data. The third reason as to why I picked this period is that it is relatively recent. Whereas many papers focus on longer periods, for example by also considering the 1990s in addition to the 2000s and 2010s in their sample, such as Pedersen et al. (2021) and Halbritter et al. (2015). The benefit of looking at a recent period is that the results will arguably be more relevant to today's stock market. For example, consider the fact that mispricing might disappear gradually, with arbitrageurs discovering the benefit of trading on ESG and therefore removing the mispricing.

Bebchuk et al. (2013) for example, who find that trading on governance, while beneficial in the 1990s, no longer generates abnormal returns in the 2000s. Therefore, even if trading on ESG in the past might have generated abnormal returns, it may not necessarily be the case anymore.

Another important choice with respect to the data and sample is the measurement of ESG itself. In this paper I choose ESG scores, following Halbritter and Dorfleitner (2015) as well as Kempf and Osthoff (2007) for example. The other possible main approach is to look at a proxy variable, such as sub industry Hong and Kacperczyk (2009), or GHG emissions as in Pedersen et al. (2021). The main advantage of using ESG scores is that it gives a broad perspective, as it weights multiple proxy variables together to have a broad coverage of firms' sustainable performance. It might also be reasonable to suggest that this is more useful to regular investors as they might not necessarily be interested in specific proxy variables as these might be both harder to track and quantify.

Finally, I used data on the market capitalization and prices of all firms from Bloomberg. The Fama & French factors, the market excess returns and the risk-free rate are all from Kenneth French’s Data Library (French, 2022).

4.2 Descriptive Statistics of ESG Scores

Below you will find a table covering the most important and basic descriptive statistics of the ESG scores which make up the basis of the sorting.

Table 1. Descriptive Statistics of ESG Scores, 2009-2019.

This table presents the mean values of the each ESG score during the entire period as well as their min and max values. It displays two different ways of calculating the standard deviation of the ESG scores, this is explained below.

ESG Scores:	E	S	G	ESG
Mean:	47.77	50.02	62.27	52.15
Max:	95	92	94	86
Min:	23	21	35	33
Std Overall:	11.17	10.74	8.80	8.34
Std Within:	2.15	1.68	1.58	1.33
T:	124	124	124	124
N:	510	510	510	510

The means of the ESG scores are all relative similar, varying between 47.77 and 62.27. There is also a relatively large discrepancy in terms of the highest and lowest ESG scores as these range from 21-35 to approximately 90. For the E scores for example, the maximum score is roughly four times as large as the minimum, with 95 and 23 respectively. In terms of the high minus low portfolios and the empirical testing, the discrepancy can be seen as a positive. It suggests that there will be a relatively large discrepancy in terms of the average ESG scores of the high and low sub portfolios. And so, if there is a general discrepancy in terms of the returns between high and low ESG stocks this sample should be sufficient to test it.

The overall standard deviation (std overall) of the ESG scores I calculated by first taking the standard deviation between the ESG scores each time t, and then taking the standard deviation of this over the entire sample period, T. The standard deviation within (std within) is calculated by first taking the standard deviation within the ESG scores of each stock over the entire sample period, and then calculating the standard deviation of these values. The latter one is arguably more interesting as it indicates how much the ESG scores vary over time within each firm. The std within is relatively small for all ESG scores in comparison with the overall means, as these are all approximately 1-2. Meaning that the ESG scores of the companies themselves do not change very much throughout the period.

5. Results and Discussion

This chapter presents and analyses the results of the paper after having applied the method and data described in the previous two chapters. First, I show the results of the CAPM, FF3 and FF5 regressions in terms of the abnormal returns, alpha. The most important result of the paper is that few of the high minus low portfolios inhibit significant alphas. The second sub section shows the betas of all regressions, indicating what factors that are correlated with each portfolio. I then proceed to show the Sharpe ratios of the sub portfolios and their mean ESG scores. Finally, the results of the mean monthly excess returns of the high minus low portfolios are presented.

5.1 High Minus Low Portfolio Abnormal Returns

The abnormal returns generated by each high minus low portfolio are presented in Table 2. It shows the alphas generated by using the CAPM, FF3 and FF5 with their respective p-values for each ESG score and sub portfolio weighting method. Starting with the high minus low portfolios with equally weighted sub portfolios, only one of these exhibits a significant alpha, namely, that of the portfolio sorted on G, when using the FF5, which has an alpha of -0.005 or -0.5% per month and a p-value of 0.029. This indicates that stocks with a relatively low G, or poor governance, tend to outperform those with good governance as defined by ESG. This is also suggested when using the FF3 as this alpha is -0.003, however this is not significant. Further, when using the CAPM this effect is seemingly non-existent with a relatively low alpha and a p-value of 0.854. Except for the portfolio sorted on G and using the FF5, the results of the regressions suggests that there are no abnormal returns, as there is a clear lack of statistical significance. Meaning that there is no discrepancy in terms of the risk adjusted returns of high and low ESG scored stocks. Further, many of the high minus low portfolios with equal weights have monthly alphas of less than 0.1% in absolute value, showing just how weak the effect of ESG is in this case.

Table 2. Monthly alpha and p-values of the high minus low portfolios.

The sample period for the equal- and value-weighted sub portfolio cover 2009-2019. The optimal portfolios cover 2012-2019 as the first three years as used to estimate the first expected returns vector and covariance matrix.

Equally-weighted sub portfolios	E	S	G	ESG
CAPM alpha:	0.003	-0.001	0.000	0.000
p-value:	0.259	0.502	0.854	0.954
FF3 alpha:	0.000	-0.002	-0.003	-0.002
p-value:	0.849	0.344	0.184	0.381
FF5 alpha:	0.000	-0.002	-0.005*	-0.002
p-value:	0.816	0.459	0.029	0.225
Value-weighted sub portfolios	**E**	**S**	**G**	**ESG**
CAPM Alpha:	0.001	-0.006*	0.001	-0.006**
p-value:	0.596	0.016	0.727	0.006
FF3 Alpha:	-0.001	-0.005*	-0.001	-0.006**
p-value:	0.785	0.026	0.692	0.004
FF5 Alpha:	-0.002	-0.006*	-0.003	-0.008***
p-value:	0.405	0.017	0.205	0.000
Optimal sub portfolios	**E**	**S**	**G**	**ESG**
CAPM Alpha:	0.005	0.002	0.002	-0.001
p-value:	0.185	0.653	0.634	0.876
FF3 Alpha:	0.004	0.002	0.000	-0.002
p-value:	0.267	0.628	0.931	0.452
FF5 Alpha:	0.004	0.002	0.000	-0.002
p-value:	0.276	0.629	0.891	0.417

Next, the high minus low portfolios using value-weighted sub portfolios generate significant negative abnormal returns when sorting on S and ESG. When sorting on S the alpha is -0.006 when using both CAPM and FF5, and -0.005 when using FF3 all of which are significant at the 5% level. The abnormal returns of the portfolio sorted on the overall ESG score are even smaller, at between -0.006 and -0.008. This indicates a discrepancy in returns between the high and low stocks. More specifically that low S and ESG stocks tend to outperform high S and ESG stocks. There appears to be no relationship between the portfolios sorted on E and G, and abnormal returns. Indicating that there is no mispricing related to these ESG pillars.

The high minus low portfolios with optimal sub portfolios do not generate any significant results. As with the other general results, suggesting a lack of mispricing related to ESG. The absence of significant abnormal returns further suggests that there is no discrepancy in the risk-adjusted returns between optimal portfolios based on high and low ESG stocks. When it comes to whether using optimal sub portfolios improved the results in terms of the abnormal returns it can be argued that it did in one way. The results of the high minus low portfolios that use optimal sub portfolios are the most consistent as they do not generate any significant alphas, for any factor or asset pricing model. It is nonetheless somewhat contradictory to what was hypothesized in chapter 3. That optimal sub portfolios may improve the performance of the high minus low portfolios as compared to equal- and value-weighted. And thereby be more likely to generate abnormal returns. This was based on the idea that portfolios formed using optimization should outperform those that are equal- and value-weighted. However, this is clearly not the case as can be observed in table 4, section 5.3. The high optimal sub portfolios tend to perform the best in terms of Sharpe ratio, whereas the equivalent low portfolios tend to perform the worst. This is likely the reason behind the complete lack of significance for the high minus low portfolios based on optimization, as opposed to the other portfolios. This will be further discussed in the section focusing on the Sharpe ratios.

The results of the FF5 regression on the portfolio sorted on G might be an anomaly as the results are not significant for the two other equivalent portfolios, or when using different asset pricing models. It does also seem somewhat contradictory as it would mean that firms with a low G score, meaning poor governance, outperform those with good governance. This lacks empirical support, Pedersen et al (2021) and Bebchuk et al. (2013), etc. all find the complete opposite results. On the other hand, the monthly mean excess returns are indeed significantly negative as can be seen in table 5, section 5.4. On the other hand, the FF5 does have stronger empirical support and is a more adequate way to test for abnormal returns than FF3 and CAPM (Fama & French, 2015). Thereby conversely suggesting a potential negative relationship with G and abnormal returns despite the findings of previous research.

The significant abnormal returns of the portfolios sorted on S and ESG using value weights is arguably the main deviation from the general results. As none of the other portfolios with an equivalent sorting consistently generate significant alphas. The result regarding S is consistent with Pedersen et al. (2021) who also find significant negative abnormal returns when sorting on S and using value weights, but not when using equal weights.

It is also consistent with the finding of a sin premium by Hong and Kacperczyk (2009) as well as Luo and Balvers (2017). It is different from the findings of Kempf and Osthoff (2007), who in contradiction find some evidence of a positive effect of S on abnormal returns. Further, it is also in contrast to Halbritter and Dorfleitner (2015) who find no significant impact of S on abnormal returns. The significant result of the overall ESG score with value weights is not consistent with Halbritter and Dorfleitner (2015) as well as Pedersen et al. (2021), both of which find insignificant results.

The relevancy of my results using value-weighted sub portfolios is to a certain extent questionable. First, because of the inconsistency with respect to the results of previous research, particularly regarding the overall ESG score. Secondly, because of the weighting method itself. The portfolios using equal and optimal weights do not generate significant alphas, suggesting that the result is driven by a few or a single stock(s). As these weighting methods limit the weight of any “extreme” stocks, particularly equal weighting. Further, as the abnormal returns persist despite controlling for a size effect when using the FF3 and FF5. It is possible that it is either underperforming stock(s) with high ESG or overperforming stock(s) with low ESG, or a combination of both. Stocks who simultaneously have a relatively high market cap, meaning that they have a relatively large weight and thereby can potentially greatly impact the results. It does not seem that this impacted the overall scores of the sub portfolios as these are relatively similar compared to the other equivalently sorted sub portfolios, as can be seen in table 7, section 5.3. And so, this effect is likely not related to ESG itself as if this was the case there should be greater deviations in the mean scores of the sub portfolios.

The most consistent result with respect to the abnormal returns is that of the portfolios sorted on E. In this case the alphas are all insignificant, for all weighting methods and asset pricing models. Clearly indicating that there is no mispricing related to the environmental impact of a firm. Meaning that investors interested in this specific ESG pillar should receive adequate risk-adjusted returns. This result is inconsistent with Pedersen et al. (2021), Derwall et al. (2005) as well as Kempf and Osthoff (2007) who tend to find significant positive abnormal returns when sorting on E. Halbritter and Dorfleitner (2015) when sorting on E tend to find no relationship between E and abnormal returns, consistent with my results.

The most interesting fact about the results is arguably the general lack of significance of the abnormal returns of the high minus low portfolios. With the primary exceptions being that of the portfolios sorted on S and ESG when using value weights. For a moment not considering the exceptions, what the results generally suggest is a lack of abnormal returns with respect to ESG. What this basically means is that there is no difference in terms of the risk adjusted returns between high and low ESG stocks. In practical terms an investor will generally not “lose out” if picking relatively high ESG stocks rather than low. As the results do not indicate a relationship between ESG and financial performance. This point should however be considered with a certain degree of caution. As the results when using value weights clearly indicate that there may be multiple outliers related to ESG. Meaning that although generally true, there seems to be some firms that go against the trend. Firms with a relatively high or low ESG scores that either over or underperform. With diversification this does not seem to be an issue, but when investing in individual companies it may be cause for concern. Further, using a zero-investment strategy with sorting based on S and ESG with value weights seems to be able to generate abnormal returns, a possible risk arbitrage strategy. If investors want to profit on ESG related mispricing the focus should however arguably lie on outlier firms. Although the zero-investment strategy works with value weights, it does not make sense to trade using the entire portfolio if the results are just driven by a few outliers. A zero-investment type strategy can instead be implemented that is focused on outliers, rather “naively” sorting a broad spectrum of stocks. If of course outliers are indeed present, even if the results indicates this it is nonetheless somewhat speculative.

I believe that the key result, particularly for ESG and its implications on investments is the results of the sorting based on overall ESG score. The overall ESG score is arguably the most important simply because it is an overall ESG score. Unless an investor has particular preferences and only wants to trade on a specific ESG pillar, such as the environmental pillar, it seems reasonable that anyone seeking to make ethical investments will most likely consider ESG as a whole. Going back to the discussion in the previous paragraphs, what the results of the high minus low portfolios with equal and optimal weights suggest is that there is no discrepancy between high and low ESG stocks. Meaning that generally investors with ESG preferences should receive a similar compensation to those disregarding it. However, as can be seen when using value weights there may still be some level of relationship between the overall ESG score and abnormal returns. Particularly on a firm-by-firm basis. In this specific aspect the results can instead be viewed as somewhat inconclusive.

With respect to the theoretical literature on equilibrium asset pricing, the general lack of significant abnormal returns suggests that ethical investing is not enough to have a substantial impact on asset pricing. With a possible exception for some individual companies, as indicated by the abnormal returns of the S and ESG scores using value weights. The relatively low std within the ESG scores indicates that companies do not act to improve their ESG performance, but on the other hand that they don't tend to decrease very much either [2]. Going back to the equilibrium model of Heinkel et al. (2001) for example, in which a relatively large amount of ESG investors can cause incentives for polluting firms to reform into nonpolluting firms. If you hold this model true to reality (which of course is a strong assumption in and of itself) and that this sample is representative of the actual US stock market. It suggests that there aren't enough investors with strong ESG preferences to have a substantial impact on stock returns and the cost of capital for firms. As this would cause incentives for firms to improve their ESG performance, and thereby scores. Further, if this was the case there would consistently be significant negative abnormal returns, particularly for the E scores. As there would be underinvestment in polluting firms, thereby causing underpricing relative to high E firms. The ESG-adjusted CAPM of Pedersen et al. (2021) have similar implications wherein a relatively large amount of ESG investors could cause a shift in the cost of capital. In terms of Merton (1987) it does not seem that low ESG stocks are neglected, and that the equilibrium expected returns of high and low ESG stocks are comparable.

A potential impact on the results are the cut-off points, as I look at the stocks with the 10% highest and lowest ESG scores respectively. It is of course possible that if the portfolio cut-off was different from 10%, say 2.5%, there would be a greater effect of ESG on the abnormal returns. However, as the cut-off point is reduced, the stocks will be less representative of the general pool of investment opportunities. It is likewise possible to suggest that a cut-off of 10% is too extreme and that it should be increased. But in this case my paper would likely generate even smaller abnormal returns as the potential effect of mispricing based on ESG should decrease even further. It is also worth noting that although the portfolios of all three weighting methods are based on the same data, the alphas generated by those with optimal sub portfolios don't cover the first three years. This is a potential point of criticism when it comes to comparing the three. Although, as previously argued, because this is a relatively consistent period it should make them comparable.

[2] A possible subject of another study.

5.2 High Minus Low Portfolio Betas

Table 3. shows the betas of the first asset pricing model, the CAPM. The most notable result is the fact that all market betas are negative. This means that the low ESG stocks are more correlated with the market, compared to the high ESG stocks. What this indicates is that low ESG stocks generally carry higher systematic risk, specifically market risk. Most of the portfolios exhibit a significant market beta, S, G and ESG when using equal and value weights as well as G when using optimal weights. As could be seen in table 2 in the previous sub section most of the abnormal returns are insignificant for the same portfolios. Meaning that in most cases risk, as modelled by the CAPM, is adequately compensated for and the assets do not deviate from the SML. Further, what the results indicate is that the high minus low portfolios with optimal weights carries the least amount of market risk. As only the portfolio sorted on G has a significant market beta. And so, even though the portfolios with optimal weights do not generate significant alphas, this suggests that this weighting method carries the least amount of market risk. Except when sorting on G of course.

Table 3. Monthly CAPM betas and p-values of the high minus low portfolios.

The sample period for the equal- and value-weighted sub portfolio cover 2009-2019. The optimal portfolios cover 2012-2019 as the first three years as used to estimate the first expected returns vector and covariance matrix.

Equal-weighted sub portfolios	E	S	G	ESG
CAPM beta:	-0.081	-0.168**	-0.422***	-0.195***
p-value:	0.168	0.002	0.000	0.000
Value-weighted sub portfolios	**E**	**S**	**G**	**ESG**
CAPM beta:	-0.004	-0.219**	-0.354***	-0.109*
p-value:	0.945	0.001	0.000	0.040
Optimal sub portfolios	**E**	**S**	**G**	**ESG**
CAPM beta:	-0.024	-0.110	-0.464***	-0.109
p-value:	0.820	0.320	0.000	0.250

Further, the G pillar has the most consistent results as the CAPM betas are significant and negative using all weighting methods. The betas themselves are relatively large, between -0.35 and -0.46 monthly, all with p-values less than 0.001. This suggests a relatively large risk exposure of low G stocks, firms with poor governance.

The increased risk of having poor governance is interestingly clearly compensated for with higher returns, as the results in the previous section show a clear lack of significant alphas. The portfolios sorted on E notably do not generate any significant beta, when using any weighting method. This indicates that there is no difference between high and low E stocks in terms of market risk. Meaning that firms who have a relatively large environmental impact do not carry more systematic risk and vice versa, as modelled by CAPM.

Next, table 4 shows the betas generated by the FF3 which in addition to the market factor also includes SMB and HML. When adding SMB and HML the market factor tends to disappear for most of the portfolios, as compared to when using the CAPM. Meaning that the market factor basically acted as some form of proxy for other effects, more adequately modelled by the FF3. The most consistent exception to this is once again the portfolios sorted on G, where the market betas are all negative and significant. Suggesting the same thing as CAPM, stocks with a poor governance are more exposed to market risk relative to those with good governance.

SMB, the factor capturing the size effect shows mixed results with respect to significance. It has a negative significant impact on most of the equal weighted portfolios and the portfolio sorted on ESG with optimal weights. Suggesting that there is some level of a size effect related to ESG. As the betas are negative it means the effect is primarily present amongst relatively low ESG stocks. The reason as to why it is present when using equal weights may be that relatively small firms have a disproportionate weight. And because small firms tend to outperform big firms, this needs to be compensated for with a higher risk, SMB beta. Interestingly this effect does not seem to be present when using value weights. This indicates that the sample does not seem to have a particular bias towards small or big firms.

The FF3 HML factor, capturing the value effect, consistently generates negative and significant betas for the portfolios sorted on G and E. It is also present when using equal and optimal weights when sorting on overall ESG. Which suggests that there is a certain degree of a value effect amongst firms with a relatively low G and E as well as overall ESG. This added risk exposure indicates that some of the low ESG stocks tend to be value stocks, stocks that have a high book-to-market value. Contrary, high ESG stocks tend to be growth stocks that instead receive investments on future promise. This is particularly interesting when sorting on E, as it may for example suggest investments in “green technology” firms. Firms that are not profitable now but rather might deliver on their promises in the future. Therefore, perhaps also related to the possibility of market incentives related to the environment.

However, this is highly speculative regarding the scope of my paper and may therefore be the subject of future research. Another interesting result of the betas of the HML factor is their consistent result with respect to G. The negative betas do in this case suggest that firms with relatively poor governance have a higher risk with respect to the HML. Indicating that firms with poor governance tend to have a high book-to-market value, perhaps related to governance itself.

Table 4. Monthly Fama-French three-factor betas and p-values of the high minus low portfolios.

The sample period for the equal- and value-weighted sub portfolio cover 2009-2019. The optimal portfolios cover 2012-2019 as the first three years as used to estimate the first expected returns vector and covariance matrix.

Equal-weighted sub portfolios	E	S	G	ESG
Market-RF:	0.018	-0.120*	-0.310***	-0.106*
p-value:	0.720	0.034	0.000	0.034
SMB:	-0.148	-0.210*	-0.198*	-0.225**
p-value:	0.061	0.020	0.039	0.005
HML:	-0.635***	0.005	-0.650***	-0.371***
p-value:	0.000	0.953	0.000	0.000
Value-weighted sub portfolios	E	S	G	ESG
Market-RF:	0.079	-0.224**	-0.285***	-0.074
p-value:	0.153	0.001	0.000	0.195
SMB:	-0.166	-0.145	-0.055	-0.145
p-value:	0.060	0.165	0.565	0.108
HML:	-0.441***	0.372***	-0.544***	-0.017
p-value:	0.000	0.000	0.000	0.833
Optimal sub portfolios	E	S	G	ESG
Market-RF:	-0.003	-0.104	-0.394***	-0.025
p-value:	0.977	0.372	0.000	0.783
SMB:	-0.054	-0.047	-0.261	-0.372**
p-value:	0.707	0.777	0.069	0.005
HML:	-0.545***	0.188	-0.838***	-0.349**
p-value:	0.000	0.215	0.000	0.004

Finally, table 5 shows the betas of the FF5, which adds the two additional factors RMW and CMA along with slightly altering both HML and SMB. The market factor is consistently present amongst the portfolios sorted on G, indicating the same results as the CAPM and FF3.

Poor governance firms are more exposed to market risk. With respect SMB the results are like those of the FF3, even when adding the additional factors. It therefore suggests similar results as previously discussed. The value effect is also still present, perhaps even more so as one additional portfolio exhibits a significant HML beta. When using the FF3 the portfolio sorted on overall ESG with value weights did not have a significant value effect, whereas it does use the FF5. This seems to be a bit of an anomaly and the results when using FF5 should be seen as a more adequate result. What this basically suggests is that there very much seems to be a value effect related to ESG, except for S in specific. Indicating that there are more value stocks amongst the firms with a relatively low ESG. This could be argued as being a sign of underinvestment in low ESG firms. However, this does not actually translate into abnormal returns as the alphas, presented in the previous subsection, do not tend to be positive.

The two additional factors, RMW and CMA, do not seem to have a major impact on the results. The CMA factor in particular only generates one significant beta, for the portfolio sorted on overall ESG with value weights. This factor is positive meaning that the effect is instead related to firms with a relatively high ESG. Because it is the only significant beta for this factor it should be considered with some degree of caution. The betas related to RMW are positive and significant when sorting on E and G, with equal and value weights. This indicates that firms with a relatively high E and G have a more robust profitability. And therefore, need an additional risk factor to compensate for this. One final note, the portfolios with optimal weights have the least amount of factor exposure and therefore seems the be the least risky weighting method.

Table 5. Monthly Fama-French five-factor betas and p-values of the high minus low portfolios.

The sample period for the equal- and value-weighted sub portfolio cover 2009-2019. The optimal portfolios cover 2012-2019 as the first three years as used to estimate the first expected returns vector and covariance matrix.

Equal-weighted sub portfolios	E	S	G	ESG
Market-RF:	0.044	-0.130*	-0.250***	-0.081
p-value:	0.377	0.026	0.000	0.110
SMB:	-0.086	-0.245*	-0.102	-0.197*
p-value:	0.287	0.010	0.278	0.018
HML:	-0.629***	0.063	-0.729***	-0.402***
p-value:	0.000	0.569	0.000	0.000
RMW:	0.298*	-0.125	0.513***	0.162
p-value:	0.016	0.377	0.000	0.190
CMA:	0.080	-0.098	0.346	0.175
p-value:	0.599	0.576	0.050	0.254
Value-weighted sub portfolios	**E**	**S**	**G**	**ESG**
Market-RF:	0.114*	-0.203**	-0.225***	-0.020
p-value:	0.035	0.003	0.000	0.717
SMB:	-0.049	-0.162	0.059	-0.104
p-value:	0.569	0.141	0.527	0.245
HML:	-0.344**	0.255	-0.606***	-0.216*
p-value:	0.001	0.050	0.000	0.041
RMW:	0.563***	-0.031	0.604***	0.241
p-value:	0.000	0.851	0.000	0.074
CMA:	-0.096	0.330	0.275	0.580**
p-value:	0.552	0.109	0.114	0.001
Optimal sub portfolios	**E**	**S**	**G**	**ESG**
Market-RF:	0.008	-0.112	-0.367**	0.001
p-value:	0.939	0.357	0.001	0.990
SMB:	0.015	-0.102	-0.199	-0.372*
p-value:	0.926	0.588	0.217	0.013
HML:	-0.597**	0.250	-0.905***	-0.373*
p-value:	0.001	0.219	0.000	0.020
RMW:	0.221	-0.165	0.234	0.038
p-value:	0.376	0.567	0.343	0.867
CMA:	0.123	-0.121	0.236	0.183
p-value:	0.670	0.718	0.409	0.482

5.3 Sharpe Ratios and Mean ESG Scores

Table 6 shows the annualized Sharpe ratios and monthly average ESG scores for each individual sub portfolio. Starting with the sub portfolios sorted on E, those with high E scores consistently outperform those of low E scores in terms of Sharpe ratio, for all weighting methods. Suggesting that forming portfolios on a relatively high E score is preferable compared to low a E score. Somewhat counter intuitive as the high minus low portfolios did not generate any significant abnormal returns, as can be seen in section 5.1. Even in the case of the relatively large difference in the Sharpe ratios of the high and low sub portfolios with optimal weights, at 0.978 and 0.577 respectively. Further, the Sharpe ratios of the high E portfolios are theoretically and empirically consistent in the sense that the optimal portfolio dominates the equal-weighted portfolio which in turn dominates the value-weighted portfolio (Bodie, Kane & Marcus, 2014). However, the results are the complete opposite for the portfolios with a low E, wherein the optimal portfolio is dominated. In terms of the mean E scores there does not seem to be any notable differences.

For the sub portfolios sorted on S there is no consistency in terms of which portfolio dominates, high or low. For the equally- and value-weighted sub portfolios those with low S scores have higher Sharpe ratios than the ones with a high S score. Whereas for the optimal sub portfolios the one with a high S dominates those with a low S. The difference in performance is relatively large for the value-weighted sub portfolios, wherein the one with low S nearly has double the Sharpe ratio of the one with high S; their Sharpe ratios being 1.072 and 0.553 respectively. This is consistent with the results from section 5.1 in terms of a significant and negative alpha for the equivalent high minus low portfolio, using all asset pricing models. Further, the mean S score does not seem to deviate very much from those of the other equivalent sub portfolios. This suggests that the significant abnormal returns are not driven by a firm, or a few firms, with a relatively extreme score, but rather with a relatively high performance, as previously discussed. Finally, as was the case with E, for the sub portfolios with high S, the optimal sub portfolio has the highest Sharpe ratio followed by the equal-weighted and then value-weighted. Whereas for the sub portfolios with a low score the optimal portfolio performs the worst.

Table 6. Annualized Sharpe ratios and average monthly ESG scores for each sub portfolio.

The sample period for the equal- and value-weighted sub portfolio cover 2009-2019. The optimal portfolios cover 2012-2019 as the first three years as used to estimate the first expected returns vector and covariance matrix.

Sharpe ratios	High	Low	Mean ESG Scores	High	Low
Equal-weighted sub portfolios:					
E	0.942	0.686	E	69.72	32.37
S	0.747	0.900	S	70.20	33.78
G	0.667	0.800	G	77.02	46.85
ESG	0.813	0.828	ESG	68.07	40.07
Value-weighted sub portfolios:					
E	0.875	0.745	E	71.02	32.68
S	0.553	1.072	S	70.04	33.03
G	0.740	0.744	G	77.60	45.06
ESG	0.592	1.118	ESG	68.11	39.47
Optimal sub portfolios:					
E	0.978	0.577	E	69.97	32.58
S	0.759	0.643	S	70.55	34.24
G	0.762	0.723	G	76.43	46.41
ESG	0.589	0.728	ESG	68.67	40.39

The sub portfolios based on a sorting of G all have relatively comparable performance in terms of Sharpe ratios. There is some discrepancy for those based on equal weighting, consistent with the abnormal returns generated by the FF5, but not by the FF3 or CAPM. In terms of the consistency of the Sharpe ratios, a similar pattern can once again be observed as with the sub portfolios sorted on E and S. For the high sub portfolio, the optimal sub portfolio has the highest Sharpe ratio, whereas it has the lowest when looking at the low sub portfolios.

Finally, the sub portfolios with stocks sorted on the overall ESG score. For the equal-weighted sub portfolios the performance is relatively similar, with Sharpe ratios of 0.813 and 0.828 respectively. However, for those based on value and optimal weights there is a relatively large difference, particularly for the sub portfolios that are value-weighted. The Sharpe ratios, in the latter case are 0.592 and 1.118 for the high and low portfolios respectively, indicating a clear outperformance. This relatively large discrepancy between the sub portfolios sorted on ESG does, like S, translate into positive abnormal returns, as shown in subsection 5.1.

With respect to consistency, the optimal sub portfolios are outperformed in all cases except one. When compared to the high value-weighted sub portfolio, in which the Sharpe ratios are relatively similar, 0.589 and 0.592.

The three examples with the biggest differences in terms of Sharpe ratio of high and low are the value-weighted sub portfolios based on S and ESG as well as the optimal sub portfolios based on E. Two of these examples translates into significant abnormal returns, the value-weighted high minus low portfolio sorted on S and overall ESG, sub section 5.1. This may suggest that the factors in the asset pricing models better captures the risk of the portfolio sorted on ESG and S rather than E. The difference in Sharpe ratio between the high and low equal-weighted sub portfolios based on G does not particularly stand out in the results, as these are 0.667 and 0.8 respectively. The high minus low portfolio in this case did however still generate a significant alpha when using the FF5. As such there is seemingly little consistency in terms of sub portfolio Sharpe ratio translating into abnormal returns.

The reason as to why a relatively large discrepancy in Sharpe ratio does not necessarily translate into abnormal returns is because the two performance measurements are simply not equivalent. Although the fundamental logic is the same behind testing for alpha and Sharpe. To evaluate performance by putting returns in relation to risk, originating in mean-variance theory. They are different ways of doing so. The asset pricing approach does of course use factors that either directly or by proxy measure risk. Whereas the Sharpe ratio considers standard deviation, which has both an idiosyncratic and systematic part. The factors therefore do not necessarily measure risk in the same way. Take the CAPM for example, which only considers systematic risk. Meaning that the difference in performance measures is related to the risk factors in the asset pricing models. And that it is not necessarily a case of incoherent performance measures. Further, the Sharpe ratios should only be seen as a complement to the main result of the paper, namely the abnormal returns. Nonetheless, the two largest differences in Sharpe ratios do indeed translate into significant abnormal returns, and so there is some degree of consistency.

One interesting result regarding the Sharpe ratios is that the optimal portfolios tend to perform the best for the high portfolios, as to be expected. Whereas the opposite is true for the optimal portfolios with low ESG scores. Amongst the low portfolios, those that are value-weighted tend to perform the best. Although the high portfolios are relatively consistent with what is to be expected from the benefits of diversification. The low portfolios seemingly completely defy this logic.

This suggests, along with the abnormal returns of the portfolio sorted on S with value weights, that low ESG stocks are prone to outliers. It seems that the portfolio optimization method did not manage to capture this effect. Simply because optimized portfolios should dominate equal-weighted portfolios which in turn should dominate value-weighted portfolios (Bodie, Kane & Marcus, 2014). Which of course generally held true for those sub portfolios with high ESG scores, but not those with low. This anomaly is a possible subject of a further study, as it clearly indicates that there may potentially be issues when executing portfolio optimization using relatively low ESG stocks. In the context of the purpose of this paper it is however not necessarily an important result. The primary purpose was to compare high and low ESG stocks, and not a test of absolute portfolio performance or portfolio optimization itself. For papers that specifically investigates portfolio optimization with ESG integration, see Pedersen et al. (2021) for example.

5.4 Mean Excess Returns of the High Minus Low Portfolios

The mean monthly excess returns of all the high minus low portfolios are shown below in table 7. The mean excess return of the portfolios sorted on E are all positive, however none are significant. Consistent with the absence of abnormal returns when sorting on E, as previously shown in section 5.1 table 2.

For the portfolios sorted on S there is the notable case of the portfolio based on value-weighted sub portfolios. The mean excess return of this portfolio is -0.9%, with a high significance level, clearly indicating that in this case the stocks with a relatively low S score generate higher returns relative to those with a high S score. This result is consistent with those of the asset pricing models, as these indicate negative and significant abnormal returns, as previously shown. The difference does not disappear even when considering risk, indicating that it is indeed due to mispricing.

The mean returns of the portfolios sorted on G are all negative. However, only the mean return of the high minus low portfolio based on equal weights is significant. This is relatively consistent as the abnormal returns generated by the FF5 are also negative and significant. However, this is not the case for the FF3 and CAPM, as these indicate that once systematic risk is considered the abnormal returns are priced away, table 2.

And so, just because there is a discrepancy between high and low G scores in terms of mean returns, it is not necessarily a case of mispricing.

Table 7. Monthly high minus low mean excess monthly returns and p-values.

The sample period for the equal- and value-weighted sub portfolio cover 2009-2019. The optimal portfolios cover 2012-2019 as the first three years as used to estimate the first expected returns vector and covariance matrix. Note that only the monthly, not excess, returns are used to evaluate the high minus low portfolios, equation 6.

Mean High Minus Low Monthly Excess returns	E	S	G	ESG
Equal-weighted sub portfolios:	0,001	-0,004	-0,006*	-0,003
p-value:	0,566	0,069	0,049	0,214
Value-weighted sub portfolios:	0,001	-0,009***	-0,004	-0,007***
p-value:	0,729	0,000	0,176	0,000
Optimal sub portfolios:	0,004	0,000	-0,004	-0,002
p-value:	0,242	0,986	0,361	0,461

The mean returns of the high minus low portfolios sorted on the overall ESG score are all negative, however only the one based on value weights is significant. As was the case with those sorted on S. This does, as with S, also translate into negative significant abnormal returns, table 2. Meaning that the discrepancy in returns does not disappear even when considering the risk factors of the asset pricing models. Indicating that the difference is indeed because of mispricing.

6. Conclusion

In summary, the results indicate that there is little to no relationship between ESG and abnormal returns. Meaning that there tends to be no difference between high and low ESG stocks in terms of risk-adjusted returns. This is because the high minus low portfolios simply tend to not generate significant alphas, particularly when using equal and optimal weights. For investors with ESG preferences this implies that they will receive similar risk-adjusted returns as those disregarding it. This rather strong conclusion should however be taken with some level of caution. Particularly when it comes to the stocks sorted on the S and overall ESG scores. As when using value weights these portfolios generate significant negative alphas. Meaning that there is some indication that stocks with a low S and ESG outperform those of a high S and ESG. It is however argued that this result is most likely driven by single or a few outliers as this effect is not present when using equal-weighted or optimal sub portfolios and despite controlling for a size effect. Meaning that investors with a diversified position will likely receive adequate compensation even when only investing in high ESG stocks. However, because of the indication that there might be outliers this is may not necessarily be the case on a stock-by-stock basis.

Implementing a similar investment strategy to the one in this paper generally would not be successful because of the absence of abnormal returns. Instead, investors wanting to take advantage of any potential mispricing related to ESG should focus on outlier firms. Which the results indicate may indeed be prevalent regarding the overall ESG and S scores. Using portfolio optimization to weight the sub portfolios did not have a substantial effect as in this case the high minus low portfolios all generated non-significant alphas. This also suggests that there is no discrepancy between the returns of optimal portfolios based on low and high ESG respectively. With respect to the theory of equilibrium pricing models; the results suggests that low ESG stocks are generally not neglect and that ethical investors are not numerous enough to substantially impact on the cost of capital and therefore the incentives of firms to reform. Something that may be investigated further and with more specification, along with the indication that outliers may be present.

An anomaly in the results is that of the portfolio sorted on G when using equal weights as this portfolio has a negative significant alpha when using the FF5. It is suggested that it is an anomaly simply because it goes against both previous research but also standard financial theory. As it would suggest that firms with poor governance outperform those with good governance, as defined by ESG.

With respect to the betas generated by the CAPM, FF3 and FF5 there are some consistent effects. Portfolios sorted on G seem to be particularly exposed to market even when considering multiple factors simultaneously. An indication that firms with poor governance carry higher systematic risk. There seems to be a value effect related to ESG, specifically for the portfolios sorted on E and G, but also to a certain extent S and overall ESG. There is little prevalence of a size effect, the betas based on SMB tends to only be significant when using equal weighting. The two additional factors of the FF5 seem to not greatly impact the results. The CMA factor in particular has little effect, whereas RMW has some effect on E and G. Fairly consistent with the value effect, suggesting that firms with high E and G have more robust profitability. Finally, the portfolios with optimal weights have the lowest exposure to the risk factors.

The results of the Sharpe ratios of the sub portfolios are somewhat mixed as they are not necessarily consistent with the abnormal returns of the high minus low portfolios. They are however different ways of measuring risk-adjusted returns. The Sharpe ratios of the individual sub portfolios still serve as an indication of how portfolios of relatively low and high ESG perform. And thereby by proxy also stocks. One inconsistency related to the Sharpe ratios is how the optimal sub portfolios perform. For the high ESG sub portfolios the optimal sub portfolios tend to perform the best, followed by equal-weighted and then value-weighted. This is consistent with both empirical findings and financial theory. For the low ESG portfolio the contrary is however true. This may be related to outliers amongst low ESG stocks as previously discussed. It also suggests that portfolio optimization may not work adequately for low ESG stocks. This may have been improved with a different optimization process, but it is also outside the scope of this paper. And it is therefore a possible subject of further study.

Finally, the mean monthly excess returns show some discrepancies between high and low ESG stocks. However, this effect mostly disappears when considering risk and it is therefore not necessarily a case of mispricing.

7. References

Alexander, G.J. (2000). On Back-Testing ”Zero-investment” strategies, *The Journal of Business*, vol. 73, no. 2, pp. 255-278.

Amel-Zadeh, A., Serafeim, G. (1991). Why and How Investors Use ESG Information: Evidence from a Global Survey, *Financial Analysts Journal*, vol. 74, no. 3, pp. 87–103.

Balvers, R.J., Luo, A.R. (2017). Social screens and systematic investor boycott risk, *Journal of Financial and Quantitative Analysis*, vol. 52, no. 1, pp. 365-399.

Bebchuk, L.A., Cohen, A., Wang, C, C.Y. (2013). The learning and disappearing association between governance and returns, *Journal of Financial Economics*, vol. 108, no. 2, pp. 323–348.

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

Best, M.J., Grauer, R.R. (1991). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results, *The Review of Financial Studies*, vol. 45, no. 2, pp. 315–342.

Black, F., Litterman, R. (1990). Asset Allocation: Combining Investor Views with Market Equilibrium, Goldman Sachs Fixed Income Research Note, September 1990.

Blitz, D., Fabozzi, F.J. (2017). Sin Stocks Revisited: Resolving the Sin Stock Anomaly, *The Journal of Portfolio Management*, vol. 44, no. 1, pp. 105–111.

Bodie, Z., Kane, A., Marcus, A. (2014). Investments, New York: McGraw-Hill-Irwin.

Bolton, P., Kacperczyk, M. (2021). Do investors care about carbon risk?, *Journal of Financial Economics*, vol. 142, no. 2, pp. 517–549.

Carhart, M.M. (1997). On Persistence in Mutual Fund Performance, *The Journal of Finance*, vol. 52, no. 1, pp. 57-82.

Chopra, V.K., Ziemba, W.T. (1993). The Effecs of Errors in Means, Variances and Covariances on Optimal Portfolio Choice, *Journal in Portfolio Management*, vol. 19, no. 2, pp. 6-11.

Danthine, J.P., Donaldson, J.B. (2014). Intermediate Financial Theory, Oxford: Academic Press.

Debondt, W.F.M., Thaler, R. (1985). Does the Stock Market Overreact?, *The Journal of Finance,* vol. 40, no. 3, pp. 793-805.

Debondt, W.F.M., Thaler, R. (1987). Further Evidence On Investor Overreaction and Stock Market Seasonality, *The Journal of Finance*, vol. 42, no. 3, pp. 557-581.

Derwall, J., Guenster, N., Bauer, R., Koedijk, Kees. (2005). The Eco-Efficiency Premium Puzzle, *Financial Analysts Journal*, vol. 61, no. 2, pp. 51-63.

Dutta, A., Bouri, E., Noor, H. (2018). Return and volatility linkages between CO2 emission and clean energy stock prices, *Energy*, vol. 164, no. 1, pp. 803-810.

Eccles, R.G., Ioannou, I., Serafeim, G. (2014). The Impact of Corporate Sustainability on Organizational Process and Performance, *Management Science*, vol. 60, no. 11, pp. 2835-2857.

Fama, E.F., French, K.R. (1993). Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics*, vol. 33, no. 1, pp. 3-56.

Fama, E.F., French, K.R. (1996). Multifactor Explanations of Asset Pricing Anomalies, *The Journal of Finance*, vol. 51, no. 1, pp. 55-84.

Fama, E.F., French, K.R. (1998). Value versus Growth: The International Evidence, *The Journal of Finance*, vol. 53, no. 6, pp. 1975-1999.

Fama, E.F., French, K.R. (2015). A five-factor asset pricing model, *Journal of Financial Economics*, vol. 116, no. 1, pp. 1-22.

Fama, E.F., Macbeth, J.D, (1973). Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy*, vol. 81, no. 3, pp. 607–636.

French, K.R. (2022). Data library, Available online: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html [Accessed 2022-05-01]

Gaunt, C. (2004). Size and Book to Market Effects and the Fama French Three Factor Asset Pricing Model: Evidence From the Australian Stock market, *Accounting and Finance*, vol. 44, no. 1, pp. 27-44.

Glushkov, D., Statman, M. (2009). The Wages of Social Responsibility, *Financial Analysts Journal*, vol. 65, no. 4, pp. 33–46.

Halbritter, G., Dorfleitner, G. (2015). The wages of social responsibility – where are they? A critical review of ESG investing, *Review of Financial Economics*, vol. 26, no. 1, pp. 25–35.

Haan, M., Dam, L., Scholtens, B. (2012). The drivers of the relationship between corporate environmental performance and stock market returns, *Journal of Sustainable Finance & Investment*, vol. 2, no. 3-4, pp. 338-375.

Heinkel, R. Kraus, A. Zechner, J. (2001). The effect of green investment on corporate behavior, *Journal of Financial and Quantitative Analysis*, vol. 36, no. 4, pp. 431–449.

Hong, H., Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets, *Journal of Financial Economics*, vol. 93, no. 1, pp. 15–36.

Jorion, P. (1986). Bayes-Stein Estimation for Portfolio Analysis, *The Journal of Financial and Quantitative Analysis*, vol. 21, no. 3, pp. 279-292.

Kim, Y., Park, M.S., Wier, B. (2012). Is earnings quality associated with corporate social responsibility? *Accounting Review*, vol. 87, no. 3, pp. 761–796.

Kempf, A., Osthoff, P. (2007). The Effect of Socially Responsible Investing on Portfolio Performance, *European Financial Management*, vol. 13, no. 5, pp. 908–922.

La Porta, R., Lakonishok, J., Shleifer, A., Vishny, R. (1997). Good News for Value Stocks: Further Evidence on Market Efficiency, *The Journal of Finance*, vol. 52, no. 2, pp. 859–874.

Markowitz, H. (1952). The Utility of Wealth, *Journal of Political Economy*, vol. 60, no. 2, pp. 151-158.

Merton, R. (1987). A simple model of capital market equilibrium with incomplete information, *Journal of Finance*, vol. 42, no. 3, pp. 483–510.

Pedersen, L.H., Fitzgibbons, S., Pomorski, L. (2021). Responsible investing: The ESG-efficient frontier, *Journal of Financial Economics*, vol. 142, no. 2, pp. 572–597.

Principles for Responsible Investment (UNPRI). (2022). Principles for Responsible Investment Annual Report 2021, Available online: https://www.unpri.org/annual-report-2021 [Accessed 2022-04-01]

Ross, S.A. (1976). The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory*, vol. 13, no. 3, pp. 341-360.

Sharpe, W.F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk, *The Journal of Finance*, vol. 19, no. 3, pp. 425-442.

Sharpe, W.F. (1966). Mutual Fund Performance, *The Journal of Business*, vol. 39, no. 1, Part 2: Supplement on Security Prices (Jan., 1966), pp. 119-138.

Seyhun, H.N. (1986). Insiders' profits, costs of trading, and market efficiency, *Journal of Financial Economics*, vol. 16, no. 2, pp. 189-212.

Sloan, R.G. (1996). Do stock prices fully reflect information in accruals and cash flow about future earnings?, *Accounting Review*, vol. 71, no. 3, pp. 289–315.

Taneja, Y.P. (2010). Revisiting Fama French Three-Factor Model In Indian Stock Market, *VISION-The Journal of Business Perspective*, vol. 14, no. 4, pp. 267-274.

Titman, S., Wei, K., Xie, F. (2004). Capital Investments and Stock Returns, *The Journal of Financial and Quantitative Analysis*, vol. 39, no. 1, pp. 677-700.

Zerbib, O.D. (2020). A Sustainable Capital Asset Pricing Model (S-CAPM): Evidence from Green Investing and Sin Stock Exclusion, Working paper.

Zhu, F. (2014). Corporate Governance and the Cost of Capital: An International Study, *International Review of Finance*, vol. 14, no. 3, pp. 393–429.