ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Specialization Financial Economics

The ESG Performance of Mutual Fund Holdings

Author: M.H. Post
Student number: 434705
Thesis supervisor: R. Barahona
Second reader: T. Eisert
Finish date: 11-09-2022

Abstract

This paper aims to research how mutual funds perform in terms of the ESG criteria of their holdings. To accomplish this data from DataStream, Morningstar, and CRSP on companies, mutual funds, and holdings is used. This paper performs multiple regression, of which several control for year- and industry-fixed effects. The results show that a higher average ESG rating of funds is correlated with a higher ESG score of companies and a higher ESG performance for three out of the four ESG performance measurements used in this paper. This paper also shows that when companies that have a relatively high average ESG rating of funds are compared with relatively low average ratings, the results do not indicate that companies with a higher average rating perform better. The results of this paper are limited by the limited use of data sources, possible selection bias, and possible omitted variable bias.

Contents

Abstract ii
List of Tables v
List of Figures v
1 Introduction 1
2 Literature Review 3
2.1 How to measure ESG 3
2.2 Mutual Fund Strategies 4
2.2.1 Screening 4
2.2.2 Using Shareholder rights 5
2.2.3 Engagement 5
2.3 Company ESG Strategies 7
2.4 Changes in ESG Criteria Companies 7
2.5 Hypotheses 8
2.5.1 First Hypothesis 8
2.5.2 Second hypothesis 8
2.5.3 Third hypothesis 9
3 Chapter 3: Data 10
3.1 Company Data 10
3.2 Fund Data 10
3.3 Holding Data 11
3.3.1 Holding Data Combined with the Morningstar Database 11
3.3.2 Holding Data combined with Datastream Data 12
4 Chapter 4: Methodology 16
4.1 ESG Performance 16
4.1.1 ESG Gain 16
4.1.2 ESG Return 16
4.1.3 ESG Relative Return 17
4.1.4 ESG Combined Returns 18
4.2 Linear- and Fixed Effect Regressions 18
4.3 ESG Score as the Dependent Variable 19
4.3.1 ESG Performance as the Dependent Variable 20
4.4 Sample Differences 20
5 Chapter 5: Results 23
5.1 Testing hypothesis 1 23
5.2 ESG Performance Measurements 24
5.3 Testing Hypothesis 2 25
5.3.1 ESG Gain 25
5.3.2 ESG Return 26

5.3.3 ESG Relative Return 26
5.3.4 ESG Combined Return 27
5.4 Testing Hypothesis 3 27
6 Conclusion 30
6.1 Limitations and Future Research 31
7 References 32

List of Tables

Table 1: Company Data DataStream 10
Table 2: Fund Data Morningstar 11
Table 3: Holding Data 12
Table 4: Holding Data combined with DataStream Data 14
Table 5: Holding Data Combined With DataStream Data Without Outliers 14
Table 6: ESG Score Regressions 24
Table 7: Summary Statistics ESG Performance Measurements With Outliers 24
Table 8: Summary Statistics ESG Performance Measurements Without Outliers 25
Table 9: Differences Sample With and Without Outliers 25
Table 10: ESG Performance Differences Threshold 3 28
Table 11: ESG Performance Differences Thresholds 3.5 and 2.5 29
Table 12: ESG Performance Differences Thresholds 3.75 and 2.25 29
Table 13: Regression Results Dependent Variable ESG Gain 26
Table 14: Regressions Results Dependent Variable ESG Return 26
Table 15: Regression Results Dependent Variable ESG Relative Return 27
Table 16: Regression Results Dependent Variable ESG Combined Return 27

List of Figures

Figure 1: Histogram ESG Score 14
Figure 2: Histogram Average Rating Funds with Outliers 15
Figure 3: Histogram Average Rating Funds without Outliers 15

1 Introduction

One of the biggest trends nowadays is environmental, social, and governance (ESG) investing. According to Reuters (2021) a record sum of $649 million billion was poured into ESG-focussed funds worldwide in 2021. Which was an increase compared to the $52 billion and $285 billion in 2020 and 2019 respectively. ESG focused funds now account for 10% of the worldwide fund assets.

This paper contributes to the current literature by showing that there is evidence that mutual funds with a high ESG rating improve the ESG criteria of their holdings more than mutual funds with a low ESG rating. A high ESG rating of mutual funds is positively correlated with the ESG score of companies. For three out of the four ESG performance measurements, there is statistical evidence that a higher average ESG rating of mutual funds is correlated with better ESG performance of companies. These results hold when controlled for year- and industry-fixed effects. When companies with a relatively high and a relatively low average ESG rating of funds are compared, there is no evidence that funds with a relatively high ESG rating improve the ESG criteria of their holdings more. A possible reason for this could be the differences in sample size.

The reasons to take ESG criteria into account when investing can differ widely. Some may care about the moral or ethical aspect of ESG criteria, while others might want to profit from a hot market. However as Michael Martin (2020) stated in the Financial Times ESG is a trend we cannot afford to ignore anymore. This is backed up by research done by the Economist Intelligence Unit (2020) where 76% of the younger generation (born in 1965-2000) in the United Kingdom say it becomes more important to consider ESG factors when investing compared to 37% of the older generation (born in and before 1964). This shows that ESG investing is socially relevant at the moment.

An easy war for private investors to take ESG criteria into account while investing is via mutual funds. According to an article written by Jon Hale for Morningstar (2021), the net flows for sustainable funds in the United States continue to attract record flows from investors. He showed that flows increased almost tenfold from $5.2 billion in 2018 to $51.1 billion in 2020. The last quarter of 2020 was also the fifth subsequent quarter to show a growth of the so-called ESG funds with a total flow of $20 billion. These sustainable funds are comprised of many different types of funds of which (open-end) mutual funds are the second biggest behind ETFs. Mutual funds are good for around 34% of the funds and are therefore an important vehicle to invest in while taking ESG criteria into account and will probably become more important as time moves on.

Due to the increase in popularity of ESG investing this subject is becoming more popular in scientific research as well. Often this research focuses on the relationship between ESG criteria and economic returns. In a review of more than 200 earlier empirical studies Friede, Busch, and Bassen (2015) find evidence for a positive relationship between ESG and Corporate Financial Performance. This is, according to them, contrary to popular belief. As many people seem to think that economic performance suffers under better ESG criteria.

There are multiple strategies possible when investing socially responsible. Wagemans, Koppen, and Mol (2013) identify three main strategies: Screening, The use of shareholder rights, and Engagement. Screening consists of the process of taking into account specified criteria that companies suitable for investment must meet (positive screening) or must not meet (negative screening). Using shareholders' rights in this context consists of using voting rights or filing proposals in shareholder meetings to pressure companies to improve ESG criteria. The last strategy engagement involves encouraging companies to improve ESG criteria informally behind closed doors. These strategies do not exclude one another and are often used in combination with each other.

Two of the three strategies discussed in the previous paragraph aim to improve the ESG criteria of companies and as ESG factors become more important for investors (The Economist Intelligence Unit, 2020), it is reasonable to assume that improving ESG criteria is also important for investors. Broccardo, Hart, and Zingales (2022) try to determine which strategies are socially optimal. Under multiple simplifying assumptions they create a theoretical framework that argues that under circumstances engagement is desirable. This, combined with the fact that mutual funds are an important vehicle for investors to invest while keeping ESG criteria in mind, leads to the following research question:

'How do mutual funds perform in terms of how their holdings improve on ESG criteria?'

This paper answers this question by first looking at the existing literature on the subject. With the knowledge of the existing literature, three hypotheses are formulated. The subsequent chapter presents the data sources and their characteristics. Chapter 4 explains the methodology. The penultimate chapter presents the results of the methodologies and discusses them. Finally, the outcome of this research and its limitations are discussed in the final chapter.

2 Literature Review

2.1 How to measure ESG

As explained in the introduction a person that wants to invest socially responsible can use environmental, social, and governance (ESG) criteria. There is no set in stone way to measure ESG criteria. Therefore this section will focus on discussing the previous literature regarding measuring ESG criteria.

Some literature focusses on one specific issue regarding ESG and often they only take negative cases in consideration. Shi, Connelly, and Sanders (2016) for instance researched whether a bigger pay gap between the CEOs and the non-CEO top of a company increased the likelihood of security class action lawsuits. Another example of research focussing on specific ESG issues is the paper of Zavyalova et al (2012). They researched if product recalls of United States toy companies influence the media coverage a company gets. These two papers give examples on what kind of ESG issues can play a role. However, their method of micro-analysing singular ESG issues is not applicable in this paper, because it would be an unrealistic amount of work to analyse multiple micro-issues to get a conclusion on macro level which the research question of this paper demands. Luckily there are different methods.

A different approach is using scores or ratings provided by data providers. Surroca, Tribo and Zahra (2013) used corporate social responsibility scores conducted by Sustainalytics, to research how multinational enterprises respond to their shareholders expectations regarding social responsibility. Similarly Kölbel, Busch and Jansco (2017) used scores provided by Asset4 based on economic, governance, environmental, and social dimensions. They use these scores to research the relation between corporate social responsibility and financial risk. Both these papers show that scores based on ESG factors can be used to research differences in ESG criteria. This method as more suited to the research in this paper than the method discussed in the previous paragraph, as it is more easily applicable on a large number of companies and/or funds.

Unfortunately there are also problems with the ESG scores. Chatterji et al (2015) research to what extent ratings of different renowned rating companies coincide. They come to the conclusion that there is a surprising lack of agreement between the rating of companies. They control for differences in the theoretical construction of the different ratings and even after that they do not find much overlap between the ratings. In line with this, Delmas and Blass (2010) also found that different methodologies in rating calculations lead to different ratings. On a slight different note Semenova and Hassel (2015) split environmental performance and environmental risk of ratings and they did find similarities between some rating companies but not between all of them. To summarize ESG scores/ratings are the easiest to use but the differences between different rating companies is food for thought about the validity of

ESG ratings. The next section will focus on what strategies mutual funds have to incorporate ESG criteria in their investment strategy.

2.2 Mutual Fund Strategies

As mentioned in the introduction there are multiple strategies possible when investing socially responsible. Wagemans, Koppen, and Mol (2013) identify three main strategies: screening, the use of shareholders rights and engagement. Screening is the process of taking specified criteria into account when making invest decisions. Companies suitable for investment must meet (positive screening) or must not meet (negative screening) these criteria. Using shareholders rights in this context consists of using voting rights or filing proposals in shareholder meetings to pressure companies to improve ESG criteria. The last strategy, engagement, involves encouraging companies to improve ESG criteria informally behind closed doors. These strategies do not exclude one another and are often used in combination with each other. The following sections will focus on the existing literature regarding these strategies.

2.2.1 Screening

As explained in the previous section there are two forms of screening positive and negative. If a fund only invests in companies that are above a certain threshold it is called positive screening, while if a fund has certain criteria's that exclude companies from investments it is called negative screening. Regarding socially responsible investing the screening phenomenon could manifest itself by differences in ESG criteria of the holdings of different mutual funds.

Curtis, Fisch and Robertson (2021) research whether ESG funds deliver on their promise. They define ESG funds in two ways. They screen mutual funds based on ESG-related keywords. They also use a list compiled by Morningstar in 2020. Curtis, Fisch, and Robertson (2021) show that on average an investor in an ESG fund is investing in an portfolio that is more tilted towards companies with a high ESG score than a non-ESG fund investor. This is in line with the screening strategy that Wagemans, Koppen, and Mol (2013) describe.

Other research involving screening often looks at the relationship between screening and fund performance. The research regarding this topic is a bit divided. Humphrey and Lee (2011) find no evidence for a difference in performance between SRI and conventional funds in Australia. Cortez, Silva and Areal (2008) show that European socially responsible funds present a performance that is comparable with conventional funds. In the contrary Wang, Liao and Zhang (2021) show that ESG screening undermines the portfolio value of funds in China. A reasons for this could be that China is a emerging market and therefore behaves differently than, for example, Europe and the United States.

The relationship between screening and firm performance is not completely clear. It is plausible that firms apply a screening strategy. Based upon the research of Curtis, Fisch, and Robertson (2021) it seems likely that firms are successfully screening on ESG criteria.

2.2.2 Using Shareholder rights

The second strategy described by Wagemans, Koppen, and Mol (2013) to improve ESG criteria is using shareholder rights. Examples of using shareholders rights are: using voting rights or filing proposals in shareholder meetings. This is a formal way of improving ESG criteria.

Dikoli et al (2021) researched shareholder proposals for Russel 3000 firms. They identified each shareholder proposal as ESG or non-ESG, after this they have a sample of 3,777 ESG shareholder proposals left. Dikoli et al (2021) also identified funds as ESG or non-ESG based upon data from Morningstar. They then researched whether ESG funds are more likely to vote in favour of ESG shareholder proposals. Dikoli et al (2021) found that on average ESG mutual funds are more likely to vote in favour of ESG shareholder proposals than non-ESG mutual funds. This research confirms that firms use their shareholder rights to try to improve ESG criteria with firms.

Similarly, the article of Curtis, Fisch, and Robertson (2021), which was also discussed in the screening section, analyses the differences in voting behaviour between ESG and non-ESG funds as well. They also show a difference in voting behaviour between the two, although they have issues pinpointing in what way they vote differently. A point that comes forward in both the article of Curtis, Fisch, and Robertson (2021) and of Dikoli et al (2021) is that fund families also can influence or explain the voting behaviour. They both account for this by adding family fund fixed effects to their regressions.

There seems to be evidence for funds using shareholder rights to increase the ESG criteria of their holding. There are at least differences in the voting behaviour of funds, although it is not completely clear what determines these differences.

2.2.3 Engagement

The last strategy that funds can use when investing socially responsible is the engagement strategy. This strategy involves encouraging companies to improve ESG criteria informally behind closed doors. This theory was described by Wagemans, Koppen, and Mol (2013). It is a more informal strategy than the previously described use of shareholder rights.

In a different paper Wagemans, Koppen, and Mol (2018). They researched this with Dutch pension funds with a multi-year survey. They found that in 2016 82% of Dutch pension funds practised

engagement while this was 33% in 2007. Their paper shows that engagement is a strategy that is being applied and is becoming more popular.

Dimson, Karakas, and Li (2015) researched several questions regarding engagement. Their dataset contains all engagement sequences for one of the largest institutional investors worldwide, measured in assets under management. In a period of ten years this large institutional investor was involved in 2,152 sequences of which 18% was successful. Although one should be cautious when extrapolating the findings regarding this single institution, it at least shows that the engagement strategy is applied by at least some firms/institutions/funds and that it can be successful. Dimson, Karakas, and Li (2015) also show that a successful engagement results in a significant positive one-year size-adjusted abnormal return of +7.1%. Perhaps more importantly they show no evidence for an adverse reaction when an engagement is non-successful, this would suggest that the probability of a non-successful engagement does not discourage attempting an engagement. These results ,that should be treated cautiously, do suggest some evidence for engagement strategies.

Broccardo, Hart, and Zingales (2022) create a theoretical framework to research what strategies that aim to impact corporate outcomes in the presence of externalities. Their theoretical frame is created under several simplifying assumptions. They claim that engagement is rarely seen, while it is theoretically optimal. A Possible explanations for this according to Broccardo, Hart, and Zingales (2022) is that possibilities to successfully engage companies is limited when somebody owns a majority of the votes. Another explanations they argue is that people have a unfounded fear for the U.S. proxy system, which investors believe limits their ability to influence corporate policy.

This paper focusses on engagement considering ESG factors. The engagement strategy can be applied on a wide variety of criteria not just ESG. Brav et al (2008); Schneider and Ryan (2009) document that hedge funds have a large heterogeneity in strategies and tactics regarding engagement. These strategies include but or not limited to informal strategies, i.e. engagement. Although both these articles are not about engagement on ESG criteria, they do show that engagement strategies are applied. There are no obvious reasons why ESG engagement should be different than other forms of engagement and therefore these results are in line with the work of Wagemans, Koppen, and Mol (2013).

There is some evidence for the engagement strategy based upon ESG criteria. This evidence is hard to extrapolate. The literature shows evidence for the engagement strategy based upon other criteria. This literature combined makes it likely that there are engagement strategies used based upon ESG criteria.

2.3 Company ESG Strategies

In the previous section, the strategies for mutual funds on how to invest socially responsible were discussed. Two of those strategies focus on improving the ESG criteria of their holdings. There are many ways in which a company can improve its ESG criteria. Therefore most literature focuses on factors that indirectly influence ESG criteria. To get an idea of what kind of actions companies can take to improve their ESG criteria, this section will give an overview of the existing literature considering this subject.

The research of Iliev and Roth (2020) show that companies, who report that they have board members that have previously served on other boards where they were exposed to changes in social and environmental changes, perform better than other firms in term of ESG ratings. In other words it beneficial for companies to have board members that experience with ESG. Another thing that companies could consider regarding their boards/management is increasing the diversity. McGuiness, Vieito, and Wang (2017) show that increasing gender diversity in the management are associated with better ESG ratings at Chinees firms.

The literature shows that there are different possibilities for companies that want to improve their ESG criteria. It could be that there are factors that limit the possibilities for companies to improve their ESG criteria. The next section will focus on this problem.

2.4 Changes in ESG Criteria Companies

When analysing changes in ESG criteria it is important to consider a few things. Exogenous shocks can seriously harm or benefit the opportunities to increase or decrease ESG criteria. For example, the Coronavirus pandemic can limit a firm's opportunities to improve governance when everyone has to work from home. These kinds of exogenous effects can influence ESG criteria and should therefore be accounted for in this paper.

Chatterji and Toffel (2009) researched how firms respond to receiving poor ratings. They find that receiving poor ratings generally leads to an improvement in performance in this field. Similar to what was argued in the previous paragraph Chatterji and Toffel (2009) control for time-varying effects by adding year fixed-effects.

Besides time-varying effects, exogenous shocks can affect companies differently depending on the industries in which companies are active. For example, an increase in oil prices may seriously affect companies in the automotive industry and therefore severely limits their possibilities to improve ESG. While an increase in oil price barely affects clothing manufacturers. For this reason, Peloza (2009) argues that it may be smart to do research in one specific sector.

The literature shows that it is wise to consider time- and industry effects when researching ESG criteria. How this will be accounted for is further discussed in Chapter 4.

2.5 Hypotheses

Now that the current literature is explained and discussed this section focuses on the hypotheses that are formed based on this research. This section tries to explain the reasoning behind every hypothesis, on which literature it is based, and what the alternative hypothesis is.

2.5.1 First Hypothesis

The first hypothesis is based upon the screening strategy described by Wagemans, Koppen, and Mol (2013). The screening strategy states that mutual funds have criteria on which they base their decision to invest, divest, or not invest in certain assets. The first hypothesis states:

H_1: The relationship between the ESG rating of a mutual fund and the ESG scores of their holdings is positive.

H_{1a}: There is no relationship between the ESG rating of a mutual fund and the ESG score of their holdings.

This hypothesis is expected to be true by definition, but it is still important to check if it holds. The research of Chatterji et al (2015); Delmas and Blass (2010); Semenova and Hassel (2015) shows that ESG scores provided by different companies can differ. Because this paper uses data from different sources, it is important to test this hypothesis. This hypothesis is in line with the literature Curtis, Fisch and Robertson (2021), as they show that investors that invest in ESG funds on average invest in a portfolio with a better ESG score. If the results of this paper are in line with their results the first hypothesis should be true.

2.5.2 Second hypothesis

The second hypothesis is based upon the engagement- and the use of shareholder rights strategies described by Wagemans, Koppen, and Mol (2013). Both these strategies assume that mutual funds actively try to increase the ESG criteria of their holdings. The second hypothesis states:

H_2: There is a positive relationship between the ESG rating of mutual funds and the ESG performance of their holdings.

H_{2a}: There is no relationship between the ESG rating of mutual funds and the ESG performance of their holdings.

If mutual funds follow engagement and/or use of shareholder rights strategies, as described by Brav et al (2008); Broccardo, Hart, and Zingales(2022); Curtis, Fisch, and Robertson (2021); Dikoli et al (2021); Dimson, Karakas, and Li (2015); Schneider and Ryan (2009); Wagemans, Koppen, and Mol (2018), then it would be logical that these strategies are more or less effective depending on whether a mutual fund has a higher or lower ESG rating. The logical relation would be that funds that have a higher rating also more successfully encourage their holdings to improve their ESG scores.

2.5.3 Third hypothesis

The third hypothesis is based upon the engagement- and the use of shareholder rights strategies described by Wagemans, Koppen, and Mol (2013). Both these strategies assume that mutual funds actively try to improve the ESG criteria of their holding. The third hypothesis is :

H_3: The holdings of mutual funds with a relatively high ESG rating on average have a higher ESG performance than the holdings of mutual funds with a relatively low ESG rating

H_{3a}: The holdings of mutual funds with a relatively high ESG rating on average do not have a higher ESG performance than the holdings of mutual funds with a relatively low ESG rating.

This hypothesis assumes that the level at which mutual funds can improve the ESG criteria depends on their ESG rating. It does not assume a causal relation. It only assumes a difference between relatively high ESG ratings and relatively low ESG ratings. Several papers found evidence for engagement strategies and for using shareholder rights strategies (Brav et al , 2008; Broccardo, Hart, and Zingales, 2022; Curtis, Fisch, and Robertson, 2021; Dikoli et al, 2021; Dimson, Karakas, and Li, 2015; Schneider and Ryan, 2009; Wagemans, Koppen, and Mol, 2018). However, not much research has been done regarding the determinants of performance on ESG criteria. Nonetheless, it is logical that funds that have a higher ESG rating are more fruitful in their attempts to increase the ESG score of their holdings than mutual funds that have a low ESG rating.

3 Chapter 3: Data

3.1 Company Data

The first data used in this paper is acquired from DataStream. The data contains information on non-financial companies in the years 2018 up and until 2022. The most important variable is the Thomson Reuter ESG score. The ESG scores are a measurement for ESG criteria that is based upon 178 data points, which result in a score from zero to a hundred. A score of zero is perceived as bad and a score of a hundred is perceived as good. Besides ESG scores and the corresponding years, this dataset also contains information on the industries in which companies are active.

Table 1 shows the summary statistics of the data. Only companies that have an ESG score for all the years in the dataset are included in the sample. Theoretically the ESG scores could range from 0 to 100. Table 1 shows that this is almost the case for the data in this paper, as the scores range from 0 to 100. The average ESG score is 45.83 with a standard deviation of 20.00. This means that the ESG scores are fairly spread out. In total the companies are active in 138 different industries. Because the total dataset contains four ESG scores for every company (one for each year) the total amount of companies in the sample is $\frac{20,635}{5} = 4,127$.

Table 1: Company Data DataStream

Variable	Mean	Standard Deviation	Min	Max	Observations
ESG Score	44.57	20.10	0.34	95.73	20,635
Industry	62.89	41.28	1	138	20,635
Year	2020	1.41	2018	2022	20,635

Notes: all number are rounded to two decimals.

3.2 Fund Data

The second database that is used in this paper contains information on mutual funds. This data is retrieved from the Morningstar Database. This data contains information on the Morningstar Sustainability Rating™. This is a rating constructed by Morningstar that gives an indication of the ESG criteria of a fund. Only mutual funds that have data on the Morningstar Sustainability Rating™ for the years 2019 up and until 2022 and are sold in the United States are part of the dataset. Table 2 shows the summary statistics of this data.

The Morningstar Sustainability Rating™ is calculated monthly in contrast to the ESG score for companies which was calculated yearly. To match the company dataset the Morningstar Sustainability Rating™ from December of the previous year was chosen to create a yearly dataset, i.e. the Rating for any fund in 2020 is equal to the rating of that fund in December 2019. This was done to get the most recent data without suffering from reverse causality. Because this paper researches the relationship of mutual funds ESG ratings on company ESG scores. It is important that the relationship is not the other

way around, i.e. reverse causality. Therefore the mutual fund ratings must be from before the ESG companies and preferable as close as possible to be as accurate as possible. This is the reason the ratings from December in the previous year are chosen.

The Morningstar Sustainability Rating™ is a rating from one to five. A low rating means a "bad" score, while a high score is perceived as "good" in terms of the ESG criteria of their holdings. In Table 2 can be seen that the average rating of a fund is 2.97. Because this dataset contains yearly data for four years the total amount of mutual funds in this dataset is $\frac{54{,}072}{4} = 13{,}518$.

Table 2: Fund Data Morningstar

Variable	Mean	Standard Deviation	Min	Max	Observations
Rating	2.97	1.01	1	5	54,072
Year	20.5	1.12	19	22	54,072

Notes: all numbers are rounded to two decimals

3.3 Holding Data

3.3.1 Holding Data Combined with the Morningstar Database

The third database enables to combine the two earlier database. It is holding data from CRSP extracted from Wharton Research Data Services. This contains data on the holdings of the mutual funds from the Morningstar Database. Because the company ESG scores are calculated at the beginning of the year. The holding data from CRSP only contains data on the first quarter of every year. To clarify, this database contains information about the investments of the mutual funds. It shows all companies that a mutual fund has investments in the first quarter of a year. Only data from funds that are in the Morningstar Database were extracted from the CRSP database. After retrieving this data all data points that were not about companies in the DataStream database were removed. After this 3,846,646 observations remain of which the summary statistics can be seen in Table 3.

Two new variables are added to Table 3 which are 'Company ID' and 'Fund ID', these are simply variables to show how many different companies and funds there are in this database. This shows that this database contains data on 3,969 companies and 3,879 funds. While only data on the same funds from the Morningstar Database was extracted from the CRSP database, the number of funds dropped from 13,518 to 3,879 a decrease of 71,3%. This has two reasons. The first reason is that some funds did not invest in companies of which ESG scores are available in the Datastream Database. This accounts for a drop to 8,788 (34.99%) funds The second reason is that Morningstar and CRSP identify funds differently. Morningstar splits up funds into sub-parts compared to CRSP. For example, Morningstar has three different identifiers for Auxier Focus funds whereas CRSP only has one identifier that is identical for all three sub-parts. Fortunately this does not lead to problems. Because although

Morningstar has different identifiers for these funds in no case these sub-parts have different Morningstar Sustainability Rating™. Therefore, the Morningstar Sustainability Rating™ can be added to the CRSP database to the identifiers that contain all the sub-parts from the Morningstar Database. This is the cause of the rest of the decline in funds to a final number of 3,879.

The amount of companies has also dropped if we compare the Datastream Database to the Holding Data in Table 3. This is because not all companies were invested in by mutual funds from the Morningstar Database. This leads to a drop from 4,127 to 3,969, which is a drop of 4.56%. The mean rating of the funds dropped from 2.97 to 2.75.

Table 3: Holding Data

Variable	Mean	Standard Deviation	Min	Max	Observations
Rating	2.75	0.92	1	5	3,846,646
Company ID	1,553	1,058.25	1	3,969	3,846,646
Fund ID	2168.26	1,114.80	1	3,879	3,846,646
Year	20.50	1.11	19	22	3,846,646

Notes: all numbers are rounded to two decimals.

3.3.2 Holding Data combined with Datastream Data

After getting the holding data this dataset was combined with the Datastream data. This was done to combine the data on the Morningstar Sustainability Ratings™ and the Thomson Reuters ESG scores for companies. After combing these datasets two new variables were calculated.

The first variable is '#Funds' which is equal to the number of funds that invest in a company in a particular year. For example, if 10 mutual funds invest in company x in 2019. The #Funds for company x in 2019 is equal to 10. One thing that should be noted is that this variable is not equal to the actual amount of funds that have been invested into a company in a year. It is only equal to the number of funds that have invested in a company in a year in which there was a Morningstar Sustainability Rating™ available. Table 4 shows that the average of companies invested is 257.65 and it also shows that the minimum of companies that have invested is 1 and the maximum is equal to 2,036. The minimum is not equal to zero because all those observations were already dropped.

The second variable that was calculated is the 'Average Rating Funds' variable. This variable is equal to the average Morningstar Sustainability Rating™ of the funds that invested in a company in a specific year. As a formula, it looks like this:

$$Average\ Rating\ Funds_{i,t} = \frac{1}{\#Funds_{i,t}} \sum_{k=1}^{\#Funds_{i,t}} Morningstar\ Sustainability\ Rating™_{k,t}$$

In this formula $\#Funds_{i,t}$ is equal to the number of companies that have a Morningstar Sustainability Rating™ and have invested in a company 'i' in year 't' as described earlier. The $Morningstar\ Sustainability\ Rating™_{k,t}$ is the Morningstar Sustainability Rating™ of fund 'k' in year 't'. Therefore, $Average\ Rating\ Funds_{i,t}$ is equal to the average rating of the funds that have invested in company 'i' in year 't'. Table 4 shows that mean of the Average Rating Funds is equal to 2.58.

Table 4 also shows that the total number of industries in which the companies are active is unchanged, compared to the DataStream data. The mean ESG score is 46.68. This is a slight difference from the DataStream sample, however, it is only a difference of 0.59, which is on a scale from 0 to 100 not economically significant. This should therefore not raise concerns about selection bias.

Figures 1 and 2 show the statistical distribution of the ESG scores and the Average Rating Funds. Both distributions look approximately normal. The ESG Score might suffer from some kurtosis but this should not be a problem. Figure 2 shows that the Average Rating Funds variable has some outliers. A couple at the low end and a few at the high end. An explanation for these outliers is that some companies only have a few funds that have been invested into them because it is easier to get more extreme values when there are less invested companies. This research assumes that the average rating of funds that is calculated using Morningstar is representative of the 'true' average ESG rating of all funds that invest in a company. For this reason, every company that has a $\#Funds_{i,t}$ of five or fewer is removed from the sample. Figures 2 and 3 show the differences in distribution this removal of data makes. Based on figures 2 and 3 it looks like the data is a bit smoother and that most outliers disappear. Especially the outliers in the upper end were caused by a low $\#Funds_{i,t}$.

Table 5 shows the summary statistics of the combined datasets after the removal of the outliers. The mean ESG score has increased slightly, which means it is closer to the original dataset from DataStream. The Average Rating Funds has also increased slightly. The mean and the minimum of the #Funds variable have both increased, as expected. The amount of observations has also decreased but there are still enough observations left to perform a good research. Chapter 5 discusses further if the removal of the outliers have consequences for the results.

Table 4: Holding Data combined with DataStream Data

Variable	Mean	Standard Deviation	Min	Max	Observations
ESG Score	45.16	19.97	.34	95.73	19,845
Average Rating Funds	2.58	0.44	1	5	14,930
#Funds	257.65	206.31	1	2036	14,930
Year	2020	1.41	2018	2022	19,845
Industry	62.35	41.03	1	138	19,845

Notes: all numbers are rounded to two decimals

Table 5: Holding Data Combined With DataStream Data Without Outliers

Variable	Mean	Standard Deviation	Min	Max	Observations
ESG Score	45.39	19.95	0.34	95.73	18,611
Average Rating Funds	2.60	0.42	1	4.13	14,642
#Funds	262.65	205.19	6	2036	14,642
Year	2019.99	1.433	2018	2022	18,611
Industry	64.36	31.03	1	138	18,611

Notes: all numbers are rounded to two decimals

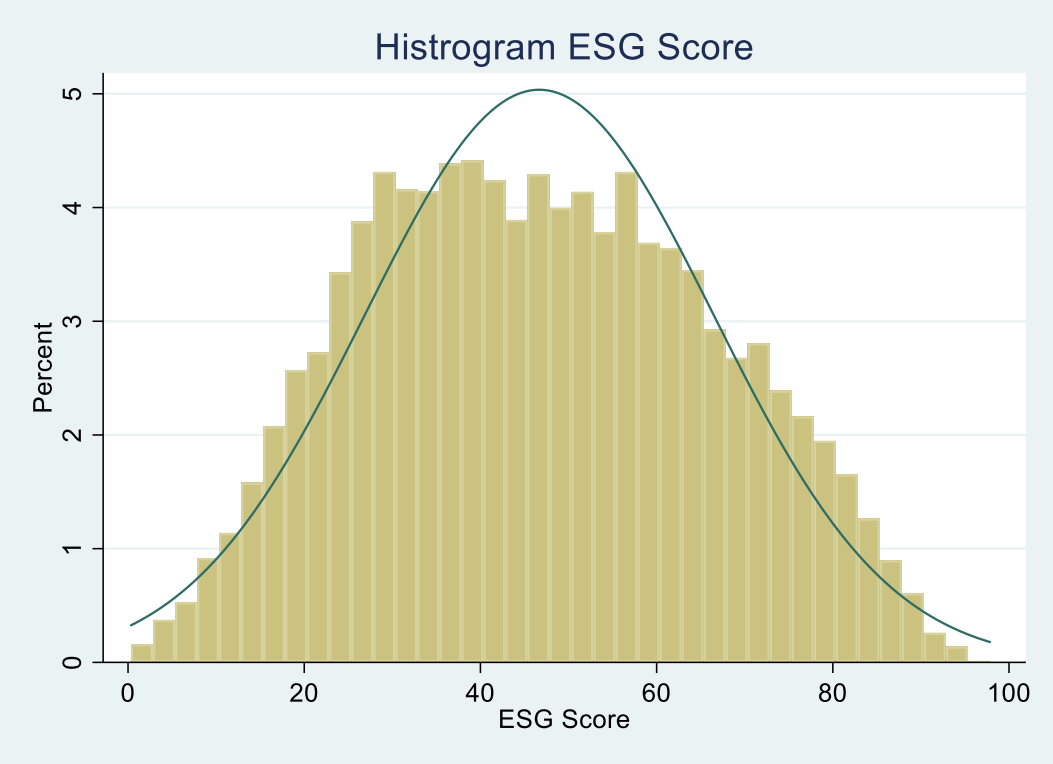

Figure 1: Histogram ESG Score

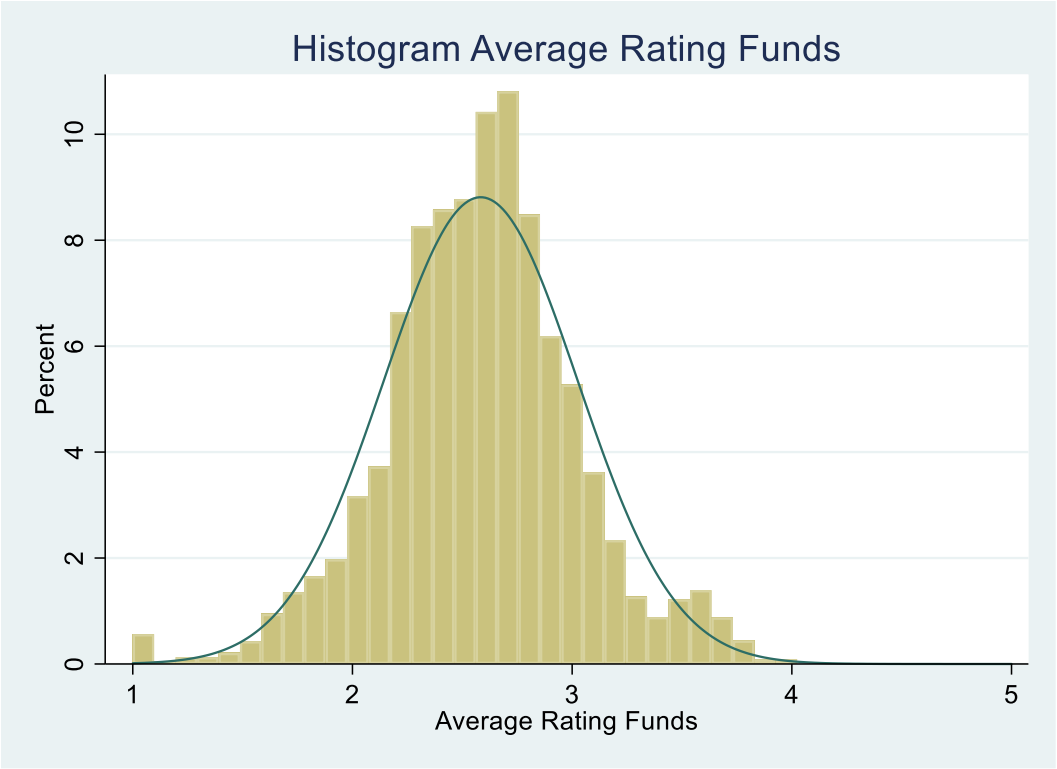

Figure 2: Histogram Average Rating Funds with Outliers

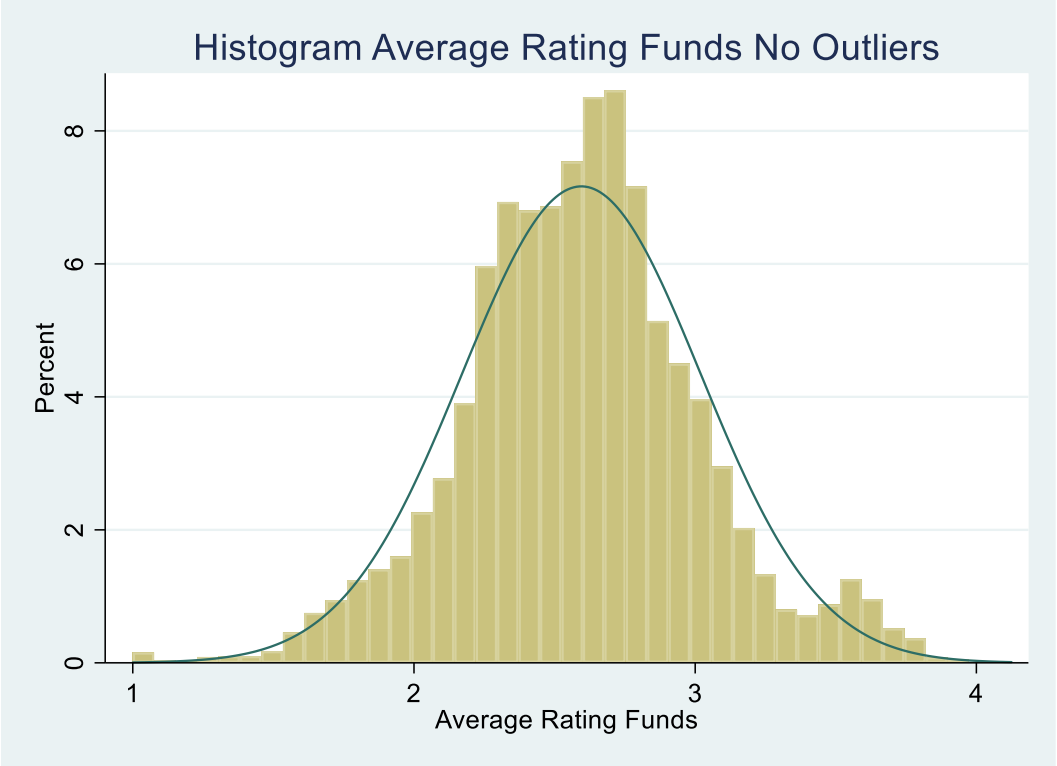

Figure 3: Histogram Average Rating Funds without Outliers

4 Chapter 4: Methodology

4.1 ESG Performance

To answer the research question and test the hypotheses it is important to have a good definition of ESG performance for companies. The tool that indicates the quality of the ESG criteria of companies in this paper is the ESG scores from DataStream, described in the Chapter 3. As described in Chapter 2 the ESG scores have limitations. Nevertheless they are the best option there is for this research paper. These ESG Scores are used to calculate four different measurements of ESG performance for companies. This section explains these ESG performance measurements, the underlying assumptions, and how the calculations.

4.1.1 ESG Gain

The first ESG performance measurement used in this paper is called ESG Gain. ESG Gain is the most straightforward measurement of the four. The formula used to calculate this measurement is:

$$ESG\ Gain_{i,t} = ESG\ Score_{i,t} - ESG\ Score_{i,t-1}$$

In this formula the ESG Gain of company 'i' in year 't' is calculated by subtracting the ESG score of company 'i' in the previous year (t-1) from the current ESG score (from year t). This measurement is quite a standard measurements. It assumes that ESG performance is constant regardless of the level of ESG score. For example, an increase from 10 to 15 in ESG score results in a ESG gain of 5. Similarly, an increase from 90 to 95 in ESG score also results in an ESG gain of 5. The assumption that ESG performance is constant regardless of the level of ESG scores is reasonable, but other assumptions are possible. The measurements discussed in the next sections have different assumptions regarding the importance of the impact of ESG scores on ESG performance.

4.1.2 ESG Return

The second ESG performance measurement used in this paper is called ESG Return. ESG Return gives a percentage-based measurement of ESG performance. The formula used to calculate this measurement is:

$$ESG\ Return_{i,t} = \frac{ESG\ Score_{i,t}}{ESG\ Score_{i,t-1}} - 1$$

In this formula the ESG Return of company 'i' in year 't' is equal to ESG score of company 'i' divided by the ESG score of company 'i' in the previous year 't-1' minus 1 (or minus 100%). Contrary to the ESG Gain measurement, explained in the previous section, this measurement does not assume that ESG performance is constant regardless of the ESG score. For example, an increase from 10 to 15 in ESG score results in an ESG return of 50% . While an increase of 90 to 95 results in an ESG return of 5.56%.

ESG Return assumes that an increase in ESG score is more impactful when the ESG score was low in the previous year than when the ESG score was high in the previous year. It also assumes that a decrease is more impactful when the ESG score was relatively low in the previous year compared to when it was relatively high.

The assumptions of ESG returns could be reasonable. However, it is also possible that it is easier to increase the ESG score when the ESG score was low. If this is correct than it contradicts the assumptions of this performance measure. Because an increase from a low ESG score would by design result in a relatively high performance. The assumptions could be true, but either way it is still useful to look at the results with this measurement. The next performance measurement has different assumptions regarding these concepts.

Another problem of this measurement is that it has a downwards limit but no upwards limit. What this means is that if the ESG Return is negative the lowest value it can get is -100%. While if the ESG Return is positive it can exceed 100%. The last performance counters this problem.

4.1.3 ESG Relative Return

The third ESG performance measurement used in this paper is called ESG Relative Return. Similar to ESG Return, ESG Relative Return also gives a percentage-based return measurement. The formula used to calculate this measurement is:

$$ESG\ Relative\ Return_{i,t} = \frac{ESG\ Score_{i,t} - ESG\ Score_{i,t-1}}{100 - ESG\ Score_{i,t-1}}$$

In this formula, the ESG Relative Return of company 'i' in year 't' is equal to the ESG score of company 'i' in year 't' minus the ESG score of the same company in the previous year. This is then divided by 100 minus the ESG Score of the previous year. The idea behind this formula is that the increase in ESG score is divided by how much improvement in ESG score was possible. For example, when a company's ESG Score increases from 10 to 15 they have an increase in ESG Score of 5, while in total an increase of 100-10=90 was possible. The ESG Relative Return is than equal to $\frac{5}{90} = 5.56\%$. Another example when a company's ESG score increases from 90 to 95 their ESG score improved by 5, while a total increase of $100 - 90 = 10$ was possible. This makes the ESG Relative Return then equal to $\frac{5}{10} = 50\%$.

The formula yields opposite results from the ESG Return measurement explained in the previous section, this is because of the assumptions of the ESG Relative Return. The ESG Relative Return assumes that it is more impactful to increase an ESG score if it was relatively high, compared to relatively low. ESG Relative Return also assumes a decrease is more impactful when the ESG score is

relatively high. These assumptions are the opposite of those of the ESG Return. For this reason, it has opposite results. The assumptions of the ESG Relative could be reasonable because it seems plausible that the marginal costs of increasing ESG score are decreasing. If this gets recognised by ESG funds then the assumptions would hold.

The ESG Relative Return has a similar problem as the ESG Return. It has a limit upwards, but not downwards. What this means is that if the ESG Relative is positive, it can have a maximum value of 100%. While if the ESG score is negative, it can have a lower value than -100%. The next performance measurement tries to counter this problem.

4.1.4 ESG Combined Returns

The last ESG performance measurement used in this paper is ESG Combined Return. It combines the ESG Return and the ESG Relative Return. The formula used to calculate this measurement is:

$$ESG\ Combined\ Returns_{i,t} = \begin{cases} ESG\ Relative\ Return_{i,t} & ,if\ ESG\ Gain_{i,t} > 0 \\ ESG\ Return_{i,t} & ,if\ otherwise \end{cases}$$

The ESG Combined Return combines the ESG Return and the ESG Relative Return, explained earlier, it therefore also combines their assumptions. The ESG Combined Return assumes that it is more impactful to increase the ESG score when the ESG score is relatively high, compared to when it was relatively low. The ESG Combined Return also assumes that a decrease in ESG score is more impactful when the ESG score is relatively low, compared to when it is relatively high. These could be reasonable assumptions. An advantage that the ESG Combined Return has over the ESG Return and the ESG Relative Return is that is both an upwards- and downwards limit. The ESG Combined Return has a theoretical range of -100% to 100%. This probably makes ESG Combined Return a more symmetrical distribution than the ESG Return and ESG Relative Return. If it is more symmetrical distributed it is less likely to cause problems with regressions.

4.2 Linear- and Fixed Effect Regressions

Now that it is clear how ESG performance is measured in this paper, it is possible to explain all the regressions used in this paper. Most of them have the ESG performance measurements as the dependent variables, but a couple also have ESG score as the dependent variable. This section explains all regressions used to test the hypotheses.

4.3 ESG Score as the Dependent Variable

The first linear regression aims to test the first hypothesis. As a reminder, the first hypothesis is:

H_1: The relationship between the ESG rating of a mutual fund and the ESG scores of their holdings is positive.

H_{1a}: There is no relationship between the ESG rating of a mutual fund and the ESG score of their holdings.

To test the first hypothesis a linear regression will be performed with the ESG score of companies as the dependent variable. This hypothesis is expected to be true by definition, but as this paper uses different data sources it is still important to test if this hypothesis holds. The linear regression will have this form:

$$(1)\ ESG\ Score_{i,t} = \alpha + \beta_1 * Average\ Rating\ Fund_{j,t} + \varepsilon_{i,t}$$

In this regression ESG score of company 'i' in year 't' is the dependent variable while the Average Rating Fund of fund 'j' in the same year is the independent variable. The alpha is the intercept and the beta is the regression coefficient, both are estimated by Stata. The epsilon is equal to the company and year-specific error term. The variable of interest in this regression is the beta. This is however not the only tool to test the first hypothesis. Several fixed-effect regressions will be performed as well. Because Chatterji and Toffel (2009) argue for using time-fixed effects, they are added to regression 2. Similarly, Peloza (2009) argues that because of industry-specific effects it is better to do research in one sector. Another option to tackle this problem is adding industry-fixed effects to the regression, this is done in regression 3. Lastly, both effects are added to regression 4. This leads to the following regressions:

$$(2)\ ESG\ Score_{i,t} = \alpha + \beta_1 * Average\ Rating\ Fund_{j,t} + \beta_2 * 2019 \ldots . + \beta_5 * 2022 + \varepsilon_{i,t}$$

$$(3)\ ESG\ Score_{i,t} = \alpha + \beta_1 * Average\ Rating\ Fund_{j,t} + \beta_2 * Industry_1 \ldots . + \beta_{139} * Industry_{138} + \varepsilon_{i,t}$$

$$(4)\ ESG\ Score_{i,t} = \alpha + \beta_1 * Average\ Rating\ Fund_{j,t} + \beta_2 * 2019 \ldots . + \beta_5 * 2022 + \beta_6 * Industry_1 \ldots . + \beta_{143} * Industry_{138} + \varepsilon_{i,t}$$

Regression (2) has added time-fixed-effects, regression (3) has added industry-fixed-effects, and regression (4) has both types of effects added. In all regressions the point of interest is β_1, this is equal

to the estimated relation between the ESG score and the Average Rating Fund. If the results are in line with the hypothesis this coefficient should be bigger than zero and statistically significant. If this is the case than the different data sources for the ESG score and ratings should not be a problem.

4.3.1 ESG Performance as the Dependent Variable

The next regressions aim to test the second hypothesis. As a reminder the second hypothesis was:

H_2: There is a positive relationship between the ESG rating of mutual funds and the ESG performance of their holdings.

H_{2a}: There is no relationship between the ESG rating of mutual funds and the ESG performance of their holdings.

To test this hypothesis multiple regressions are performed using the different measurements for ESG performance as the dependent variable. Similar to the previous regressions both time- and industry-fixed effects are added to the regressions. The additions of these fixed effects resulted in four regressions per performance measurement, this leads to sixteen extra regressions. For convenience not all sixteen regressions are typed out. Only the most complex model is shown with ESG performance measurement as the dependent variable, which is equal to one of the four measurements described earlier. This gives the following formula

$$(5)\ ESG\ Performance\ Measurement_{i,t} = \alpha + \beta_1 * Average\ Rating\ Fund_{j,t} + \beta_2 * 2019 \ldots. + \beta_5 * 2022 + \beta_6 * Industry_1 \ldots. + \beta_{143} * Industry_{138} + \varepsilon_{i,t}$$

In these regressions the ESG Performance Measurement can be one of the four measurements described in the beginning of this chapter. As a reminder all the measurements are ESG Gain, ESG Return, ESG Relative Return, and ESG Combined Return. The regressions has time- and industry fixed-effects. The point of interest in the regressions is β_1. If the results are in line with the hypothesis than this coefficient should be bigger than zero and statistically significant.

4.4 Sample Differences

The last method of this chapter focuses on testing the third hypothesis which was:

H_3: The holdings of mutual funds with a relatively high ESG rating on average have a higher ESG performance than the holdings of mutual funds with a relatively low ESG rating

H_{3a}: The holdings of mutual funds with a relatively high ESG rating on average do not have a higher ESG performance than the holdings of mutual funds with a relatively low ESG rating.

To test this hypothesis, this paper will compare the ESG performances of multiple samples. Because the definition of a 'high' and a 'low' ESG mutual fund rating is subjective, this paper uses different thresholds to define which mutual funds have a high ESG rating and a low ESG rating. The mutual fund ESG ratings can have a value of one up to five, which means that the rating that should be considered 'normal' or 'neutral' is equal to three. This is why the first samples have a threshold of three. This method compares the ESG performance measures of companies with an average mutual fund rating greater than three to those with an average mutual fund rating lower than three. The other thresholds are > 3.5 & <2.5, > 3,75 & < 2.25. The significance of the difference in ESG performance between samples is calculated using two-sample t-tests.

$$(1)\ t\text{-}stat = \frac{(\bar{x}_a - \bar{x}_b) - (\mu_a - \mu_b)}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$$

$$(2)\ t\text{-}stat = \frac{(\bar{x}_a - \bar{x}_b) - (\mu_a - \mu_b)}{s_p\sqrt{\frac{1}{n_a} + \frac{1}{n_b}}}$$

In these formulas, $\bar{x}_a$ is the average of the ESG performance measure of sample a and $\bar{x}_b$ is equal to the average of the ESG performance measure of sample b. N_a is equal to the amount of observation in sample a and n_b is equal to the amount of observations in sample b. S_a is equal to the standard deviation of sample a, S_b is equal to the standard deviation of sample b, and S_p is equal to the pooled standard deviation. The formula for the pooled standard deviation is:

$$s_p = \sqrt{\frac{(n_a - 1) * s_a^2 + (n_b - 1) * s_b^2}{n_a + n_b - 2}}$$

The variables μ_a and μ_b are equal to the expected averages under the null hypothesis of the tests. In this case the null- and alternative hypothesis are:

$$H_0: (\mu_a - \mu_b) = 0$$

$$H_b: (\mu_a - \mu_b) \neq 0$$

This means that under the null hypothesis, the difference between the average of sample 'a' and sample 'b' is non-existent and thus is equal to zero. Formula (2) assumes that variances are equal, while formula

(1) does not assume this. An F-test calculates whether the two samples' variances are statistically different. This F-test looks like this:

$$F = \frac{s_L^2}{s_S^2}$$

In this formula s_L is equal to the largest of the two standard deviations and s_S is equal to the smallest deviations of the two. The null- and alternative hypothesis of this test are:

$$H_0: \sigma_a^2 = \sigma_b^2$$

$$H_b: \sigma_a^2 \neq \sigma_b^2,$$

The calculated F-statistics are compared to critical values of an F-distribution table at a 5% significance level. If an F-statistic is greater than its critical value, then the variances are assumed to be not equal, and formula (1) is used. If the calculated F-statistic is not greater than its critical value, then formula (2) is used.

5 Chapter 5: Results

This Chapter focuses on the results and will explain what they mean. The results will test the hypotheses described in Chapter 2 and the methodology behind the results as described in Chapter 4. This Chapter tests the three hypotheses in order. Firstly, this Chapter will do multiple regressions with the ESG score of companies as the dependent variable. Secondly, This Chapter calculates and analyses the ESG performance measures explained in the previous Chapter. Thirdly, the results of the regressions with the ESG performance measures as the dependent variable are presented and explained. Lastly, this Chapter calculates, shows, and discusses the differences in ESG performance between samples.

5.1 Testing hypothesis 1

This section tests the first hypothesis. As a reminder the first hypothesis is:

H_1: The relationship between the ESG rating of a mutual fund and the ESG scores of their holdings is positive.

H_{1a}: There is no relationship between the ESG rating of a mutual fund and the ESG score of their holdings.

Four regressions test this hypothesis. Table 5 shows the results of these regressions. Table 6 shows that the coefficient for Average Rating Funds is positive and statistically significant across all the regressions. Regression 1 has a relatively high and statistically significant constant term, indicating that the fixed-effect regressions are justified. Table 6 shows that the ESG score of a company increases when the Average Rating of Funds increases, ceteris paribus. Regression 4 controls for year- and industry-fixed effects. Regression 4 shows that when the Average Rating Fund increases by 1, the ESG score of a company increases by 11.83 . These results align with the first hypothesis because the coefficient for Average Rating Funds is positive and statically significant at a 1% level in all regressions. There is statistical evidence that the relationship between the ESG rating of a mutual fund and the ESG score of their holdings is positive. This means that the use of different databases for the ESG scores for companies and funds is unlikely to cause problems.

Table 6: ESG Score Regressions

	Dependent Variable: ESG Score			
Variable	**(1)**	**(2)**	**(3)**	**(4)**
Average Rating Funds	7.16***	10.86***	7.64***	11.83***
	(.51)	(.59)	(.50)	(.57)
Constant	28.26***	NA		
	(1.36)		NA	NA
Year FE	No	Yes	No	Yes
Industry FE	No	No	Yes	Yes
Observations	14,642	14,846	14,846	14,846

Notes: all numbers are rounded to two decimals. The numbers between brackets are the standard errors clustered by firm. The asterisks indicate significance * 10%, ** 5%, and *** 1%.

5.2 ESG Performance Measurements

This section shows the results of the ESG performance calculations. Chapter 4 shows the math and reasoning behind all the ESG performance measurements. Table 7 and Table 8 show the results of the calculations. As explained in Chapter 3, some outliers were removed. Table 7 shows the summary statistics of the ESG performance measurements with the outliers, while Table 8 shows the same without outliers. The differences are almost non-existent. For completeness, Table 9 shows the statistical significance of the differences between the results of Table 7 and Table 8. The statistical significance is calculated using a two-sample t-test, as explained in Chapter 4. Table 9 shows no statistical significance for the differences between the two samples. For this reason, the remaining methods use the sample without outliers.

Table 8 shows that the averages of all ESG performance measurements are positive. The standard deviations of all sizes are relatively high compared to the means, but because of the large number of observations, the means are statistically different from zero. The minimum is negative for all measurements. As explained in Chapter 4, the ESG Return and the ESG Relative Return have a relatively high maximum and minimum, respectively. As described in Chapter 4, the underlying assumptions change when considering a positive ESG Performance measurement against a negative ESG performance measurement. Because the ESG Combined Return is a combination of these two measurements, no such extremes are seen with this measurement.

Table 7: Summary Statistics ESG Performance Measurements With Outliers

Variable	Mean	Standard Deviation	Min	Max	Observations
ESG Gain	2.55	6.77	-75.47	54.66	14,930
ESG Return	0.11	1.01	-0.92	113.94	14,930
ESG Relative Return	0.04	0.14	-4.23	0.77	14,930
ESG Combined Return	0.04	0.13	-0.92	0.77	14,930

Notes: all numbers are rounded to two decimals.

Table 8: Summary Statistics ESG Performance Measurements Without Outliers

Variable	Mean	Standard Deviation	Min	Max	Observations
ESG Gain	2.56	6.76	-75.47	54.66	14,642
ESG Return	0.11	1.02	-0.92	113.94	14,642
ESG Relative Return	0.04	0.14	-4.23	0.77	14,642
ESG Combined Return	0.04	0.13	-0.92	0.77	14,642

Notes: all numbers are rounded to two decimals.

Table 9: Differences Sample With and Without Outliers

Variable	Mean Original Sample	Mean Without Outliers	Difference
ESG Gain	2.55	2.56	7.58e-3
ESG Return	0.11	0.11	1.32e-4
ESG Relative Return	0.04	0.04	2.46e-4
ESG Combined Return	0.04	0.04	4.62e-4
Observations	14,930	14,642	

Notes: all numbers are rounded to two decimals. If a number has is followed by ex it means that you have to multiply the number by 10^x to get the true value. The asterisks indicate significance * 10%, ** 5%, and *** 1%.

5.3 Testing Hypothesis 2

This section focusses on testing the second hypothesis. As a reminder the second hypothesis was:

H_2: There is a positive relationship between the ESG rating of mutual funds and the ESG performance of their holdings.

H_{2a}: There is no relationship between the ESG rating of mutual funds and the ESG performance of their holdings.

To test the hypothesis, this section shows the results of several regressions with ESG performance measurements as the dependent variables. The first regression for every measurement only has the average rating funds as an independent variable. The second regressions have year-fixed-effects, the third regressions have industry-fixed-effects, and the last regressions have year- and industry-fixed effects. For a full explanation of the regressions, see Chapter 4.

5.3.1 ESG Gain

Table 10 shows the results of the regressions with ESG gain as the dependent variable. Table 10 shows that the coefficient for the Average Rating Funds in all regressions is significant and positive. The first regression shows an average increase of 0.39 rating on average for every point that the average rating for funds increases, ceteris paribus. The magnitude of the coefficients of Average Rating Funds is relatively low. Regression 4 shows an average increase of 0.50 in ESG gain for every increase of average rating funds of 1, ceteris paribus. Because the range of ESG Gain theoretically goes from -100 to 100,

and in the data used in this sample, it goes from -75.47 to 54.66, an increase of 0.50 points is relatively low. So although the results are statistically significant, their economic significance is questionable.

Table 10: Regression Results Dependent Variable ESG Gain

	Dependent Variable: ESG Gain			
Variable	**(1)**	**(2)**	**(3)**	**(4)**
Average Rating Funds	.39***	.53***	.35***	.50***
	(.13)	(.14)	(.13)	(.14)
Constant	1.54***			
	(.34)	NA	NA	NA
Year FE	No	Yes	No	Yes
Industry FE	No	No	Yes	Yes
Observations	14,642	14,642	14,642	14,642

Notes: all numbers are rounded to two decimals. The numbers in brackets are the standard errors clustered by firm. The asterisks indicate significance * 10%, ** 5%, and *** 1%.

5.3.2 ESG Return

Table 11 shows the results of the regressions with ESG Return as the dependent variable. Contrary to the previous results (Table 10), the regressions with ESG Return as the dependent variable show negative coefficients for the Average Rating Funds variable. A negative coefficient is the opposite of what is expected by the second hypothesis, but they are also not statistically significant. The results of Table 11 are not in line with the second hypothesis. An explanation could be that the assumptions of the ESG Return are not reasonable.

Table 11: Regressions Results Dependent Variable ESG Return

	Dependent Variable: ESG Return			
Variable	**(1)**	**(2)**	**(3)**	**(4)**
Average Rating Funds	-1.23e-3	-1.34e-3	-7.78e-3	-9.31e-3
	(1.16e-2)	(1.62e-2)	(9.44e-3)	(1.24e-2)
Constant	1.11e-1**			
	(2.63e-2)	NA	NA	NA
Year FE	No	Yes	No	Yes
Industry FE	No	No	Yes	Yes
Observations	14,642	14,642	14,642	14,642

Notes: all numbers are rounded to two decimals. The numbers in brackets are the standard errors clustered by firm. If a number has is followed by ex it means that you have to multiply the number by 10^x to get the true value. The asterisks indicate significance * 10%, ** 5%, and *** 1%.

5.3.3 ESG Relative Return

Table 12 shows the results of the regressions with the ESG Relative Return as the dependent variable. Similar to the results with the ESG Gain as the dependent variable, these regressions show a positive and statistically significant coefficient for the Average Rating Funds variable. The results are in line with the second hypothesis. Like Table 6, the first regression has a relatively high and significant constant term, indicating that fixed-effects regressions are probably justified. The first regression shows that the ESG Relative Return increases by 1.11% when the Average Rating Funds increase by 1, ceteris

paribus. When controlled for year- and industry-fixed-effects, regression 4 shows that the average ESG Relative Return increases by 1.42% when the Average Rating Funds increases by 1, ceteris paribus. The results of Table 15 align with the second hypothesis because the coefficient of Average Rating Funds is positive and statistically significant.

Table 12: Regression Results Dependent Variable ESG Relative Return

	Dependent Variable: ESG Relative Return			
Variable	**(1)**	**(2)**	**(3)**	**(4)**
Average Rating Funds	1.11e-2***	1.48e-2***	1.04e-2***	1.42e-2***
	(2.59e-3)	(2.66e-2)	(2.69e-3)	(2.75e-3)
Constant	1.21e-2*			
	(6.77e-3)	NA	NA	NA
Year FE	No	Yes	No	Yes
Industry FE	No	No	Yes	Yes
Observations	14,642	14,642	14,642	14,642

Notes: all numbers are rounded to two decimals. The numbers in brackets are the standard errors clustered by firm. If a number has is followed by ex it means that you have to multiply the number by 10^x to get the true value. The asterisks indicate significance * 10%, ** 5%, and *** 1%.

5.3.4 ESG Combined Return

Lastly, Table 13 shows the results of the regressions with the ESG Combined Return as the dependent variable. Similar to the regressions with the ESG Gain and the ESG Relative Return as the dependent variable, these regressions also all have a positive and statistically significant coefficient for the Average Rating Funds variable. This means that the results in Table 13 align with the second hypothesis. The first regressions show that an increase of 1 in the Average Rating Funds variable is associated with a rise in ESG Combined Return of 2.08%, ceteris paribus. When controlled for year- and industry-fixed-effects, regression 4 shows that an increase of 1 in Average rating funds is associated with an increase of 2.94%, all else equal. The results of Table 16 align with the second hypothesis.

Table 13: Regression Results Dependent Variable ESG Combined Return

	Dependent Variable: ESG Combined Return			
Variable	**(1)**	**(2)**	**(3)**	**(4)**
Average Rating Funds	2.08e-2***	2.90-2***	2.05e-2***	2.94e-2***
	(2.70e-3)	(2.90e-3)	(2.78e-3)	(2.98e-3)
Constant	-1.47e-2**			
	(7.18e-3)	NA	NA	NA
Year FE	No	Yes	No	Yes
Industry FE	No	No	Yes	Yes
Observations	14,642	14,642	14,642	14,642

Notes: all numbers are rounded to two decimals. The numbers in brackets are the standard errors clustered by firm. If a number has is followed by ex it means that you have to multiply the number by 10^x to get the true value. The asterisks indicate significance * 10%, ** 5%, and *** 1%.

5.4 Testing Hypothesis 3

This section focuses on answering the third hypothesis. As a reminder the third hypothesis was:

H_3: The holdings of mutual funds with a relatively high ESG rating on average have a higher ESG performance than the holdings of mutual funds with a relatively low ESG rating

H_{3a}: The holdings of mutual funds with a relatively high ESG rating on average do not have a higher ESG performance than the holdings of mutual funds with a relatively low ESG rating.

To test this hypothesis, the results in this chapter compare the ESG performance measurements of different samples. Table 14 shows the differences in ESG performance measurements of companies with a higher average fund rating than three and those with a lower average fund rating than three. The means of all the performance measurements are positive, but the differences are not. If the results align with the third hypothesis, the differences should be positive and statistically significant. The differences between the ESG gains and ESG returns are negative, which is the opposite of what the third hypothesis expects. The ESG Combined Return is the only statistically significant difference and is also bigger than zero. Therefore, the difference in ESG Combined Return is in line with the third hypothesis. It should be said that the threshold of three is quite a broad definition of high and low values.

Table 14: ESG Performance Differences Threshold 3

Average Rating Funds	**>3**	**<3**	
Variable	**Mean**	**Mean**	**Difference**
ESG Gains	2.46	2.57	-0.10
ESG Returns	9.00%	11.08%	-2.08%
ESG Relative Return	4.38%	4.05%	0.33%
ESG Combined Return	4.90%	3.76%	1.13%***
Observations	2,155	12,448	

Notes: all numbers are rounded to two decimals. The asterisks indicate significance * 10%, ** 5%, and *** 1%.

Table 15 shows the differences in ESG performance measurements of companies, with a threshold of 3.5 for relatively high and 2.5 for relatively low average fund ratings. Table 15 shows one more statistically significant difference than Table 14, the ESG Gain difference. The difference in the ESG Gains is negative and therefore not in line with the third hypothesis. The ESG combined return difference is still significant, but the significance level drops compared to the difference in Table 14. The difference in the ESG combined return is still positive, which is in line with the third hypothesis.

Table 15: ESG Performance Differences Thresholds 3.5 and 2.5

Average Rating Funds	>3.5	<2.5	
Variable	Mean	Mean	Difference
ESG Gains	1.56	2.20	-0.64**
ESG Returns	6.85%	9.95%	-3.10%
ESG Relative Return	3.23%	3.36%	0.14%
ESG Combined Return	3.95%	2.85%	1.10%*
Observations	501	5,879	

Notes: all numbers are rounded to two decimals. The asterisks indicate significance * 10%, ** 5%, and *** 1%.

Table 16 shows the last results of this section. Table 16 shows the differences between the samples, with a threshold of 3.75 for a relatively high fund rating and a threshold of 2.25 for a relatively low average fund rating. Table 16 shows more statistically significant differences than Tables 14 and 15. Three of the four differences are statistically significant. All of the differences are negative, however. Negative differences are the opposite of what the third hypothesis expects. The results are therefore not in line with the third hypothesis.

Table 16: ESG Performance Differences Thresholds 3.75 and 2.25

Average Rating Funds	>3.75	<2.25	
Variable	Mean	Mean	Difference
ESG Gains	0.03	2.06	-2.02**
ESG Returns	1.35%	9.51%	-8.16%***
ESG Relative Return	-0.74%	3.15%	-3.89%*
ESG Combined Return	1.34%	2.33%	-0.99%
Observations	66	2,665	

Notes: all numbers are rounded to two decimals. The asterisks indicate significance * 10%, ** 5%, and *** 1%.

Although Table 14 and 15 showed a statistically positive difference in the ESG combined return, the other results in Table 15 and especially Table 16 show different results. Because Table 14 has quite a broad definition of a high- and low-average fund rating, the results should be interpreted carefully. How stricter the definition for a low- and high average fund rating gets, the more statistically significant the differences become, except for the combined return. Some of the results of Table 15 and especially the results of Table 16 are not in line with the third hypothesis. Therefore, there is no statistical evidence that, on average mutual funds with a relatively high ESG rating improve the ESG scores of their holdings more than mutual funds with a relatively low ESG rating. These results contradict the precious results in Section 5.3. A possible explanation for the contradicting results is the differences in sample size. In Table 14, 15 and 16 the samples with relatively high Average Rating Funds have less observations than the samples with relatively low Average Rating Funds. This could skew the results and therefore result in a different outcome compared to the second hypothesis.

6 Conclusion

This paper aims the answer the research question:

'How do mutual funds perform in terms of how their holdings improve on ESG criteria?'

This paper tries to answer this research question by testing three hypotheses. This research uses data from DataStream on companies, Morningstar on mutual funds that are sold in the United States, and CRSP on mutual fund holdings.

The results show statistical evidence that a higher ESG rating of funds is associated with higher ESG scores in their holdings. This paper shows this by performing regressions with ESG scores as the dependent variable. Based on the literature of Chatterji and Toffel (2009) and Peloza (2009), three regressions add time- and industry-fixed-effects.. All regressions performed in this paper show a statistically significant positive coefficient for the average ESG rating of mutual funds on the ESG score of their holdings. This indicates that the use of different data bases is justified.

The results of this paper show statistical evidence that there is a positive relationship between the ESG rating of mutual funds and the ESG performance of their holdings for three of the four ESG performance measurements used in this paper. The regressions show that for ESG Gain, ESG Relative Return, and ESG Combined Return, the coefficient of the average rating of mutual funds is positive and statistically significant. Three out of the four regressions performed for every ESG performance measurement use year- and industry-fixed-effects, based on the literature of Chatterji and Toffel (2009) and Peloza (2009). The regression with ESG Return shows no statistically significant results, and this could be due to the assumptions underlying the ESG Return.

Lastly, his paper shows no statistical evidence that mutual funds with a relatively high ESG rating on average improve the ESG scores of their holdings more than mutual funds with a relatively low ESG rating. This conclusion is based upon the differences in ESG performance of companies with a relatively high average ESG rating of funds compared to a low average ESG rating of funds. How stricter the definition of relatively high and relatively low becomes, the more statistically significant the differences become, except for the ESG Combined Return. improve the ESG score of their holdings more than mutual funds with relatively high ESG ratings. These results are somewhat contradicting. A possible explanation is the differences in sample sizes between the relatively high and the relatively low samples. The results, indicate that it probably is more likely that the mutual funds with relatively low ESG ratings

All in all, it seems that mutual funds improve the ESG criteria of their holdings more when their ESG rating is high. This holds when controlled for year- and industry-fixed effects. If the relatively high and low samples are compared, an opposite effect is found, which could be to differences in sample sizes. This paper also shows that mutual funds with a high ESG rating invest in companies with a higher ESG score. This indicates that people who find ESG criteria critical when investing should probably invest in mutual funds with a higher ESG rating. This paper also shows some evidence for the strategies for socially investing described by Wagemans, Koppen, and Mol (2013).

6.1 Limitations and Future Research

One of the considerable limitations of this research is the use of ESG scores and ratings for companies and firms to indicate their ESG criteria. Chatterji et al (2015) and Delmas and Blass (2010) show that ratings provided by third parties have problems, and different parties' ratings do not always align with each other. Although this paper checks the validity of using different data bases, future research would benefit from using more databases.

Another problem with this research is that it only uses data for mutual funds that are sold in the United States. For this reason, it is not easy to extrapolate the results of this paper to mutual funds that are sold in other countries, especially emerging markets. Future research could focus on using data for mutual funds in other countries, but as of today, this data is challenging to get. A similar problem might be a selection bias in what companies and/or funds receive ratings for third parties. Future research could focus on researching whether this is the case. As time moves on and ESG becomes more popular more data should be available as well, so future research should have fewer problems with this issue.

Lastly, it could be that there are other factors that should be controlled for. The year- and fixed effects remove some of the randomness. It is, however, possible that there are other factors that influence the results, i.e. ommited variable bias. Future research could focus on identifying these factors and including them in the study.

7 References

BRAV, A., JIANG, W., PARTNOY, F., & THOMAS, R. (2008). Hedge Fund Activism, Corporate Governance, and Firm Performance. *The Journal of Finance*, *63*(4), 1729–1775. https://doi.org/10.1111/j.1540-6261.2008.01373.x

Broccardo, E., Hart, O., & Zingales, L. (in press). Exit vs. Voice. *Journal of Political Economy*. https:/doi.org/10.1086/720516

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2015). Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal*, *37*(8), 1597–1614. https://doi.org/10.1002/smj.2407

Chatterji, A., & Toffel, M. W. (2009). How Firms Respond to Being Rated. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1018719

Cortez, M. C., Silva, F., & Areal, N. (2008). The Performance of European Socially Responsible Funds. *Journal of Business Ethics*, *87*(4), 573–588. https://doi.org/10.1007/s10551-008-9959-x

Curtis, Q., Fisch, J., & Robertson, A. (2021). Do ESG Funds Deliver on Their Promises? *Michigan Law Review*, *120.3*, 393. https://doi.org/10.36644/mlr.120.3.esg

Delmas, M., & Blass, V. D. (2010). Measuring corporate environmental performance: the trade-offs of sustainability ratings. *Business Strategy and the Environment*, *19*(4), 245–260. https://doi.org/10.1002/bse.676

Dikolli, S., Frank, M. M., Guo, M. Z., & Lynch, L. J. (2021). Walk the Talk: ESG Mutual Fund Voting on Shareholder Proposals. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3849762

Dimson, E., Karakaş, O., & Li, X. (2015). Active Ownership. *Review of Financial Studies*, *28*(12), 3225–3268. https://doi.org/10.1093/rfs/hhv044

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, *5*(4), 210–233. https://doi.org/10.1080/20430795.2015.1118917

Hale, J. (2020, July 24). *Sustainable Equity Funds Are Outperforming in Bear Market*. Morningstar, Inc. Retrieved 26 April 2022, from https://www.morningstar.com/articles/972475/sustainable-equity-funds-are-outperforming-in-bear-market

Hale, J. (2021, September 27). *A Broken Record: Flows for U.S. Sustainable Funds Again Reach New Heights*. Morningstar, Inc. Retrieved 26 April 2022, from https://www.morningstar.com/articles/1019195/a-broken-record-flows-for-us-sustainable-funds-again-reach-new-heights

Humphrey, J. E., & Lee, D. D. (2011). Australian Socially Responsible Funds: Performance, Risk and Screening Intensity. *Journal of Business Ethics*, *102*(4), 519–535. https://doi.org/10.1007/s10551-011-0836-7

Iliev, P., & Roth, L. (2020). Do Directors Drive Corporate Sustainability? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3575501

Kerber, R., & Jessop, S. (2021, December 23). *Analysis: How 2021 became the year of ESG investing*. Reuters. Retrieved 26 April 2022, from https://www.reuters.com/markets/us/how-2021-became-year-esg-investing-2021-12-23/

Kölbel, J. F., Busch, T., & Jancso, L. M. (2017). How Media Coverage of Corporate Social Irresponsibility Increases Financial Risk. *Strategic Management Journal*, *38*(11), 2266–2284. https://doi.org/10.1002/smj.2647

Martin, M. (2020, November 26). *ESG: a trend we can't afford to ignore*. Financial Times. Retrieved 26 April 2022, from https://www.ft.com/content/87a922a1-8d60-4295-a9d8-d2c1ab5d788e

Peloza, J. (2009). The Challenge of Measuring Financial Impacts From Investments in Corporate Social Performance. *Journal of Management*, *35*(6), 1518–1541. https://doi.org/10.1177/0149206309335188

Schneider, M., & Ryan, L. V. (2009). A review of hedge funds and their investor activism: do they help or hurt other equity investors? *Journal of Management & Governance*, *15*(3), 349–374. https://doi.org/10.1007/s10997-009-9113-x

Semenova, N., & Hassel, L. G. (2014). On the Validity of Environmental Performance Metrics. *Journal of Business Ethics*, *132*(2), 249–258. https://doi.org/10.1007/s10551-014-2323-4

Shi, W., Connelly, B. L., & Sanders, W. G. (2015). Buying bad behavior: Tournament incentives and securities class action lawsuits. *Strategic Management Journal*, *37*(7), 1354–1378. https://doi.org/10.1002/smj.2400

Surroca, J., Tribó, J. A., & Zahra, S. A. (2013). Stakeholder Pressure on MNEs and the Transfer of Socially Irresponsible Practices to Subsidiaries. *Academy of Management Journal*, *56*(2), 549–572. https://doi.org/10.5465/amj.2010.0962

Verheyden, T., Eccles, R. G., & Feiner, A. (2016). ESG for All? The Impact of ESG Screening on Return, Risk, and Diversification. *Journal of Applied Corporate Finance*, *28*(2), 47–55.

Wagemans, F. A., Koppen, C. K. V., & Mol, A. P. (2013). The effectiveness of socially responsible investment: a review. *Journal of Integrative Environmental Sciences*, *10*(3–4), 235–252. https://doi.org/10.1080/1943815x.2013.844169

Wagemans, F. A., van Koppen, C. K., & Mol, A. P. (2018). Engagement on ESG issues by Dutch pension funds: is it reaching its full potential? *Journal of Sustainable Finance & Investment*, *8*(4), 301–322. https://doi.org/10.1080/20430795.2018.1485379

Wang, Z., Liao, K., & Zhang, Y. (2021). Does ESG Screening Enhance or Destroy Stock Portfolio Value? Evidence from China. *Emerging Markets Finance and Trade*, *58*(10), 2927–2941. https://doi.org/10.1080/1540496x.2021.2014317

Why ESG investing is on the rise (EIU). (2020). Rbcwealthmanagement. Retrieved 26 April 2022, from https://www.rbcwealthmanagement.com/en-eu/insights/why-esg-investing-is-on-the-rise-in-2020

Zavyalova, A., Pfarrer, M. D., Reger, R. K., & Shapiro, D. L. (2012). Managing the Message: The Effects of Firm Actions and Industry Spillovers on Media Coverage Following Wrongdoing. *Academy of Management Journal*, *55*(5), 1079–1101. https://doi.org/10.5465/amj.2010.0608