ESG investing in the Eurozone
Portfolio performance of best-effort and best-in-class approaches

Kajsa Andersson & Simon Mårtensson
[This page was left intentionally blank]
Acknowledgements

This thesis is written at Umeå School of Business, Economics and Statistics as our degree project.

We would like to thank our supervisor, Jörgen Hellström for his guidance and expertise throughout the process. We would also like to thank our family and friends for their support during this period.

Umeå, 2019-05-24

Kajsa Andersson & Simon Mårtensson
[This page was left intentionally blank]
Abstract
The last decades have seen a rapid increase of sustainable investing, also known as ESG (Environmental, Social and Governance) investing. There has also been an increasing body of academic literature devoted to whether investors can gain any financial benefits from taking ESG under consideration. Previous literature of portfolio performance in terms of risk-adjusted returns has given much of its attention to best-in-class approaches, which is a strategy that selects top performers in ESG within a sector or industry. The purpose of this study is foremost to investigate a best-effort approach to ESG investing, which is a strategy that focuses on the top improvers in ESG. The purpose is further to compare this with a best-in-class approach, since the findings from earlier studies of this strategy still are inconsistent. The region chosen to perform this study in is the Eurozone.

Several theories that have implications for portfolio studies and abnormal returns are taken under consideration in relation to the study and its findings. This includes the efficient market hypothesis, the adaptive market hypothesis and modern portfolio theory. The theoretical framework also cover asset-pricing models and the notions of risk-adjusted returns. A quantitative study with a deductive approach are used to form portfolios, with a Eurozone index as the investable universe. Best-effort and best-in-class portfolios as well as difference portfolios of the two approaches are created, based on ESG data and different cut-off rates for portfolio inclusion. As for risk-adjusted performance measure, the Carhart four-factor model are used.

The overall results are mostly insignificant findings in terms of abnormal returns. However, three best-effort portfolios based on the top ESG improvers show significant positive abnormal returns. These findings are strongest for the environmental and social factor. As for the best-in-class approach, only the governance portfolios provided weakly significant results in terms of abnormal returns. Further, the study is not able to significantly distinguish between a best-effort and a best-in-class approach when it comes to risk-adjusted performance. The exception is the environmental factor based on the top performers in each approach, where the best-effort portfolio outperforms the best-in-class portfolio. Finally, none of the portfolios provided significant negative risk-adjusted returns. This can at least be considered as good news for ESG investing, since it indicates that investors do not have to sacrifice risk-adjusted returns in order to invest in a more sustainable way.
Definitions

Best-in-class approach: An SRI-strategy that refers to selecting the best performing companies within a certain class or sector based on ESG score.

Best-effort approach: An SRI-strategy that refers to selecting the companies that have made the biggest sustainability development effort.

Best-in-universe approach: An SRI-strategy that refers to selecting the performing companies in the whole investable universe based on ESG score.

ESG Score: Scoring system to assess a company’s performance based on environmental concerns (such as resource use, emissions and innovation), social aspects (such as workforce, human rights, community and product responsibility), and governance aspects (such as management, shareholders and CSR strategy).

Risk-adjusted returns: An investment’s return that accounts for how much risk is involved in producing that return.

Abnormal returns: The difference between the actual return and expected return of an investment.

Abbreviations

SRI: Sustainable and Responsible Investments

ESG: Environmental, Social, Governance

CSR: Corporate Social Responsibility

Eurosif: European Sustainable Investment Forum

GISA: Global Sustainable Investment Alliance
Table of Contents

1. INTRODUCTION .................................................................................................................. 1
   1.1 Problem background ........................................................................................................ 1
   1.2 Problem discussion .......................................................................................................... 2
   1.3 Research purpose and research question ........................................................................ 5
   1.4 Theoretical and practical contribution ........................................................................... 5
   1.5 Delimitations .................................................................................................................... 6

2. SCIENTIFIC METHOD ............................................................................................................. 7
   2.1 Research Philosophy ...................................................................................................... 7
       2.1.1 Ontological considerations .................................................................................... 7
       2.1.2 Epistemological considerations ............................................................................. 8
   2.2 Research Approach ....................................................................................................... 8
   2.3 Perspective ..................................................................................................................... 9
   2.4 Choice of theoretical framework .................................................................................. 9
   2.5 Systematic literature review ......................................................................................... 9
   2.6 Source Criticism .......................................................................................................... 10

3. THEORETICAL FRAMEWORK .............................................................................................. 11
   3.1 Efficient market hypothesis .......................................................................................... 11
   3.2 Adaptive market hypothesis ......................................................................................... 12
   3.3 Modern portfolio theory .............................................................................................. 13
   3.4 Asset-pricing models and abnormal returns .................................................................. 14
       3.4.1 Capital asset pricing model (CAPM) .................................................................... 14
       3.4.2 Jensen’s alpha ........................................................................................................ 15
       3.4.3 Fama & French three-factor model ....................................................................... 15
       3.4.4 Carhart four-factor model ..................................................................................... 16

4. LITERATURE REVIEW .......................................................................................................... 18
   4.1 CSR, financial performance and value-enhancing capabilities ................................... 18
   4.2 Empirical studies on investors reactions to CSR improvements ............................... 19
   4.3 Previous studies of best-effort and best-in-class approaches .................................... 21
   4.4 Summary of previous empirical literature .................................................................. 25

5. PRACTICAL METHOD ............................................................................................................ 27
   5.1 Hypotheses .................................................................................................................... 27
   5.2 Investable universe ....................................................................................................... 28
   5.3 Data gathering .............................................................................................................. 29
   5.4 Thomson Reuters ESG score ....................................................................................... 29
   5.5 Data processing and calculations ................................................................................ 31

5.6 Portfolio construction ....................................................................................................... 32
   5.6.1 Portfolio weighting ................................................................................................... 32
   5.6.2 Cut-off rates ............................................................................................................. 33
   5.6.3 Best-in-class portfolios ............................................................................................ 33
   5.6.4 Best-effort portfolios ............................................................................................... 34
   5.6.5 Difference portfolios ............................................................................................... 34
   5.6.6 Summary of portfolios ............................................................................................ 35
   5.6.7 Transaction costs ..................................................................................................... 35

5.7 Performance measure ....................................................................................................... 35

5.8 Statistical analysis ............................................................................................................ 37
   5.8.1 Regression analysis ................................................................................................ 37
   5.8.2 Hypothesis test ....................................................................................................... 38
List of Figures
Figure 1. Deductive research method ......................................................... 9
Figure 2. Cumulative returns for the ESG portfolios .................................. 42
Figure 3. Cumulative returns for the environmental portfolios ..................... 43
Figure 4. Cumulative returns for the social portfolios ................................. 44
Figure 5. Cumulative returns for the governance portfolios ......................... 44

List of Tables
Table 1. Previous empirical literature ....................................................... 26
Table 2. Portfolios .................................................................................... 35
Table 3. ESG score and change in ESG score for all portfolios ..................... 41
Table 4. ESG score and change in ESG score for Euro STOXX 300 ............... 42
Table 5. Time-series regressions for ESG portfolios .................................. 46
Table 6. Time-series regressions for individual pillar score portfolios ............ 47
Table 7. Time-series regressions for difference portfolios ............................ 48
1. INTRODUCTION

This chapter will describe the background and discussion leading to our research question. The text will discuss both practical and academic motivation behind the choice of subject. Based on this, the research question and purpose of the thesis will be presented. Furthermore, we will discuss the potential practical and theoretical contribution as well as presenting delimitations of the thesis.

1.1 Problem background

In September 2015, leaders from all around the world gathered at the United Nations summit in New York to address the need for a more sustainable global development. The result was Agenda 2030, which consists of 17 sustainable development goals set for the year 2030. Through these aspirational goals, the UN aims to reduce poverty, hunger and inequalities, and to increase the global health and gender equality (United Nations, 2015). Further, the goals intend to produce decent work, economic growth and to provide responsible production and consumption. A couple of months after the summit in New York, the leaders of the world gathered once more. This time in Paris. The reason was to specifically address the climate change and the greenhouse gas emissions. The outcome was that 195 countries signed the Accord de Paris, or commonly known as the Paris Agreement. The target of the convention is to keep the increase in the average global temperature well below 2 °C above pre-industrial levels. Also, to increase the ability to adapt to adverse impacts of a changing climate (UNFCCC, Paris Agreement - Article 2, 2016). To meet the Paris Agreement, it has been estimated that over the next decade, Europe alone need at least €180 billion per year in additional climate investments (Eurosif, 2018, p. 6). Further, meeting the UN Sustainable Development Goals is estimated to require annual investments of between $5-7 trillion on a global scale (Eurosif, 2018, p. 71).

At a closer look on the aims and implications of these agreements, one can see that they must rely on the contribution of private corporations and investors. For example, Goal 12 of the Sustainable Development Goals is to ensure sustainable consumption and production patterns. The target of this goal is to “encourage companies, especially large and transnational companies, to adopt to sustainable practices and to integrate sustainability information into their reporting cycle” (United Nations, 2015, p. 22). Furthermore, the Paris Agreements states that one aim of the agreement is to “making finance flows consistent with a pathway towards low greenhouse gas emissions and climate-resilient development” (UNFCCC, Paris Agreement - Article 2, 2016).

So, the question going forward is if companies can adapt and take a more sustainable approach without having to sacrificing their most prominent task of maximizing profitability? From the investor’s perspective, the question is almost the same. Can investors invest in a more sustainable and social responsible way, without having to sacrifice returns?

However, even if these questions are much more present and debated in the world of business and finance today, they are hardly new. Over 50 years ago, Frederick (1960) argued that the old philosophy of business responsibility is outdated. Instead he saw the need for theories on business responsibility with a focus toward public gains, not only private gains. The last decades have seen the rise and development of the stakeholder theory, originated from the work of Freeman (1984). The theory in its essence builds on the idea that corporate social responsibility (CSR) can build strong relationships between
a firm and its various stakeholders. These stakeholders can be employees, consumers or local communities (Brooks and Oikonomou, 2018, p. 1). The relationships will then help the firm in reaching long-term success and financial strength. There is also the other side of the coin, the shareholder theory. The famous economist Milton Friedman published an article in 1970, titled “The social responsibility of business is to increase its profits”. In this article, Friedman criticized the idea that businesses should have obligations towards society and that this idea will reduce the profitability of firms, and thus shareholders’ wealth. These two opposing views has led to a question that has sought a lot of academic attention throughout the years. The question of whether a firm can do well by also doing good.

1.2 Problem discussion
The last decade has seen a rapid growth of Sustainable Responsible Investing (SRI). According to Global Sustainable Investments Alliance (GISA), the valuation of investments following SRI-principles was $22.89 trillion in the beginning of 2016, and has more than doubled since 2012 (GSIA, 2016, p. 7). Before moving on, the term SRI might need a formal definition. The European Sustainable Investment Forum (Eurosif), a member of GISA, defines SRI as:

“Sustainable and responsible investment (SRI) is a long-term oriented investment approach which integrates ESG factors in the research, analysis and selection process of securities within an investment portfolio. It combines fundamental analysis and engagement with an evaluation of ESG factors to better capture long term returns for investors, and to benefit society by influencing the behavior of companies.” (Eurosif, 2018, p. 12)

It can further be explained that ESG factors refers to environmental, social and governance aspects of a company. In other words, ESG is a way of measuring corporate social responsibility (CSR), which is a term that is commonly used by corporations and in academic research. This thesis uses the terms ESG and CSR interchangeably, as the main differences stems from which perspective one takes, the firm or the investor, or which study one refers to. Further, ESG is by nature not a clearly defined measure. There are many different sources and providers of ESG-data, and the exact calculations differs among them.

For asset managers and investors, there are several different approaches to integrate sustainable considerations into the investment decision. Eurosif (2012, p. 10) defines seven different strategies for investors:

1. Sustainable themed investment
2. Best-in-class investment selection
3. Norms-based screening
4. Exclusion of holdings from investment universe
5. Integration of ESG factors in financial analysis
6. Engagement and voting on sustainability matters
7. Impact investment

Sustainable themed investment is investments into themes or assets directly linked with sustainability issues. This can for example be in companies that provide renewable energy or clean water. The best-in-class approach refers to selecting the leader or best-
performing asset within a certain class or sector, based on the ESG score. Norm-based screening is a screening method based on companies’ fulfilment of international standards and norms. Exclusion of holdings from the investment universe is simply to exclude companies from the investable universe, for example companies active in controversial industries such as manufacturers of weapons and tobacco. Integration of ESG factors refers to include risks and opportunities of ESG and the potential impacts of these into the more traditional financial analysis. Engagement and voting are engaging in active ownership, through for example share voting on ESG-related matters. Lastly, impact investment refers to direct investment into companies, organizations or funds that seek a sustainable impact alongside with financial objectives. For example, project-specific investments in developing markets. (Eurosif, 2012, p. 10)

Eurosif’s definitions of strategies is claimed to be aligned with other counterparts in the industry of sustainable responsible investment (Eurosif, 2018, p. 13). Historically, strategies based on the exclusion of companies have been the most practiced, and still is (Eurosif, 2018, p. 16). However, in recent years, “best-in-class” and “ESG integration” has both grown tremendously. The best-in-class approach has grown from €133 billion in asset under management in 2009, to €585 billion in 2018 (Eurosif, 2018, p. 17). ESG integration on the other hand, has grown with 27% annually the last couple of years and have reached a valuation over €4 trillion (Eurosif, 2018, p. 16).

It’s clear that the integration and importance of ESG-factors and ESG-based approaches are a growing theme in the financial industry. The last decades have also seen the rise of large body of research regarding financial performance and ESG within the academic field of finance. From this research, there is studies that support the fact that investing in ESG-criteria can produce abnormal risk-adjusted returns. As for portfolio performance, Kempf and Osthoff (2007) showed that a strategy based on a best-in-class approach, where they bought the highest ranked stocks in terms of ESG within each industry and sold the lowest ranked stocks, lead to significant abnormal returns. These findings were confirmed by Statman and Glushkov (2009), who found that investors adopting a best-in-class method in their portfolio construction could outperform conventional portfolios. However, a recent study by Halbritter & Dorfleitner (2015) cover a longer time period, and concludes that the positive returns witnessed in the before mentioned studies disappears in later years. This is only a few of the many studies in terms of ESG investing and best-in-class approaches. On an overall basis, Friede et al. (2015, p. 225) concludes that portfolio-related studies based on ESG in general shows mixed findings.

Supporters of the efficient market hypothesis would argue that investment strategies based on ESG-information couldn’t lead to abnormal returns, since it’s publicly available information. Event-studies suggests that investors do not value a firm’s engagement in ESG-improving issues, as demonstrated in Hawn et al. (2018), Kappou and Oikonomou (2016), and Michlik and Rubash (2011). These studies use inclusion or exclusion from sustainable indices as proxy for news of improvement or deterioration in corporate social responsibility. Further, Krüger (2015) investigated how stock markets reacts in the short-run to both negative and positive CSR events. He showed that investor reacted negatively on negative events. Quite surprisingly, he also found that investor reacted weakly negative on positive events in most of the cases. On the other hand, previous research tends to lean towards a positive relationship between the level of CSR and overall financial performance, demonstrated in studies such as, Orlitzky et al. (2003). There is also studies that support corporate social responsibility to have several value-enhancing
capabilities for firms (Malik, 2015). Assuming that higher ESG is in fact value-enhancing for the firm, it would be viewed as an intangible asset. Studies have shown that the market are slower to adapt and recognize intangible values (Edman, 2011). To sum up, investors does not seem to value ESG improvements straight away. On the other hand, many studies document a positive relationship between ESG and firm performance and value. This would indicate that firms that are improving their ESG scores are undervalued, and that the benefits from engaging in ESG related issues would then be incorporated in the stock price over time rather than straight away.

Most of the academic findings focuses on the level of ESG and financial performance. That is, does higher rating in ESG have a relationship with higher firm performance, higher returns or lower risk. Going back to the strategies of SRI, the academic research is well aligned with the best-in-class approach, through its selection of ESG-leaders. However, according to Eurosif (2018, p. 10), the best-in-class also includes “best-in-universe” and “best-effort” approaches to ESG investing. Eurosif refers to AMF (2015, p. 10) for the definition of the best-effort approach, which states that “this approach seeks to include in the portfolio only issuers that have made the greatest sustainable development effort. Issuers that have made the most progress are not necessarily best-in-universe in ESG”. Thus, this approach focuses on the companies that have improved their effort within ESG the most, rather than the ones with highest scores. The idea with best effort in ESG or CSR in terms of portfolio performance and risk-adjusted returns is something that haven’t received much attention in the academic field. Even though some of the theoretical framework that support high ESG-levels can be applied to ESG-changes as well. As far as the authors of this thesis can tell, the only other academically oriented study with a direct focus on a best-effort approach is Benlemlih et al. (2018). The authors investigate the effects that direct changes in CSR-ratings can have on firm performance, risk and stock returns. By focusing on year-by-year changes in CSR. they find that CSR changes is significantly related to higher firm value, as well as related lower systematic risk. As for portfolio performance, the strategy significantly lowered portfolio performance in terms of risk-adjusted returns. Benlemlih et al. (2018) base their study on a sample of U.S companies during the years 1996-2011. Thus, more regions and different time periods needs academic scrutiny to assess this gap in the research.

Busch et al. (2016, p. 315) argues that investors will learn how to price ESG-information the same way as they have learned how to price financial information, at least in the long run. This would indicate that high-scoring companies already have higher valuations. Further, Busch et al (2006, p. 315) advises investors to take a forward-looking perspective on ESG-information. These considerations seem reasonable. Also, since ESG-factors is something that’s been around the financial markets for some time now, it is reasonable that investors by now have learned to price the information in current levels of ESG correctly. Thus, implementing a strategy with companies that improves their ESG scores, along with the potential financial benefits of these improvements, could provide a fruitful strategy for investors. As academic evidence point toward a positive relationship between ESG levels and financial performance, it would be reasonable to identify trends of companies with improving ESG scores, as these companies would be able to benefit from better financial performance in the future. In the problem background of this thesis we can see that the biggest sustainability challenges may still be ahead of us. Therefore, looking at firms who improves their ESG score may be in the interest for both society and investors, rather than to look at the ones that are already performing well. This would imply a best-effort for portfolio management. As mentioned, the only study found by the
authors of thesis that concerns a best-effort approach is conducted on US firms, as most of the studies concerning best-in-class approaches also are. Therefore, this study find Europe, and especially the Eurozone due to its similarity in currency, as a suitable market to further investigate a best-effort approach.

1.3 Research purpose and research question
The reasoning for a socially responsible investor to invest by ESG-criteria in general or by best-in-class or best-effort approaches in particular is clear. That is, to invest funds in companies that are in the front edge of corporate social responsibility, or in companies that are looking to improve. There are also academic argument supporting that these strategies can generate abnormal returns. Previous academic studies have focused on whether companies that score high in ESG can improve portfolio performance, with inconsistent findings. Thus, the academic focus has been on a best-in-class approach. The purpose of this thesis is primarily to investigate a best-effort approach in terms of risk-adjusted portfolio performance. Thereby, this thesis aim to investigate whether a strategy based on firms that improve their ESG-score can generate abnormal returns.

Another purpose is to compare this strategy with a best-in-class approach, since earlier research on the subject is inconsistent, and mostly conducted on firms in the U.S. To fulfil these purposes, this thesis formulates the following research questions:

- Is there a relationship between abnormal returns and a best-effort approach to ESG-investing in the Eurozone?
- Is there a relationship between abnormal returns and a best-in-class approach to ESG-investing in the Eurozone?
- Is there a difference between a best-effort and a best-in-class approach in terms of abnormal returns in the Eurozone?

1.4 Theoretical and practical contribution
This thesis hopes to contribute to the growing academic field of how corporate social responsibility affect and relates to portfolio performance and risk-adjusted returns. More specifically, by focusing mostly on changes in ESG score rather than the level of the score, this thesis aims to address this issue from another angle and perspective. Practically, this thesis will contribute on how to integrate and handle ESG data in terms of portfolio-management. The results will provide further and deeper knowledge on how to further incorporate ESG data in investment decisions. This will be of interest for investors in general, and for socially responsible investors in particular. ESG-investing is foremost seen as long-term oriented investment approach. Focusing on ESG-improvers is more of a short-term based strategy. Thus, this research could help to provide motivation for portfolio managers of SRI-funds to expand their strategies without losing focus on ESG criteria.
1.5 Delimitations

The delimitations of this thesis have been made to conduct a suitable and reliable study within the timeframe of this thesis, with the aim of fulfilling the purpose and research questions. The delimitations are listed below:

- There are several providers of ESG ratings. This study will only use ESG data from Thomson Reuters, which is available through their database Thomson Reuters Eikon.
- For practical reasons, such as similarity in currency and limitation of countries, this study will as mentioned restrict the sample to only incorporate the Eurozone. Because of practical considerations this study aims to have all returns denoted in the same currency. If other European countries would have been included in the sample, the exchange rates between euro and other currencies could have driven the returns this study aims to investigate. For example, the Nordic countries have throughout the years received high ratings in ESG (RobecoSAM, 2019). This could possible lead to a tilt towards these countries in some of our portfolios. The portfolio returns under the period could thereby be driven, to some extent, by an altering exchange rate between these countries’ currencies and the Euro.
- This study will only define the Euro STOXX 300 as an investable universe for our representative Eurozone investor. Euro STOXX 300 is an index that consist of approximately 300 Eurozone companies.
- This study will limit itself to only constructing equally-weighted portfolios, due to time constraints. Value-weighted portfolios would perhaps be preferable, since it is more commonly used by investors (Statman & Glushkov, 2009, p. 42). However, equally-weighted portfolio is used in many similar studies earlier as well, such as Benlemlih et al. (2018), Kempf & Osthoff (2007) and Statman & Glushkov (2009).
- The period in the study will only cover the years of 2009-2018 for the portfolio performance. Studying a longer period would be preferable, especially considering the relatively one-sided rise in stocks prices since the financial crisis. Ten years is nevertheless assumed to be enough to provide a reliable study.
- Only year-to-year changes will define improvements in ESG for the best-effort approach in this study. One could argue for other measurements of improvements in a best-effort approach. For example, improvements over a longer period. Year-to-year improvements is however considered the most straightforward measurement, and it is also used by Benlemlih et al (2018).
- This study will consider only the Carhart four-factor model as performance measure for risk-adjusted returns. However, there are many other measurements for risk-adjusted returns, such as CAPM, Fama & French three-factor model and the Sharpe ratio. The reason to only use Carhart four-factor model is due to time constraints. Also, the model seems to be the most commonly used measurement in other similar studies concerning ESG investing, such as Benlemlih et al. (2018), Kempf & Osthoff (2007) and Halbritter & Dorflieitner (2018).
2. SCIENTIFIC METHOD

This chapter will present the scientific method applied in this thesis. Decisions regarding research philosophy and approach will be clarified and explained. Potential influence of our authors perceptions will be presented together with the study’s ontological and epistemological considerations. This chapter also covers source criticism, research perspective and choice of theoretical framework.

2.1 Research Philosophy

The researcher should consider the development of knowledge and the nature of the research reality (Bryman & Bell, 2015 p. 5). That is, one’s research philosophy, which often tends to influence the whole research process and how results are interpreted and constructed. Assumptions and beliefs regarding research philosophy, and their relation to theory, are divided into paradigms of two. Ontology which concerns the perception of the reality (Bryman & Bell, 2015 p. 26). Epistemology which concerns the theory of knowledge (Bryman & Bell, 2015 p. 32). Perceptions of what knowledge is, how it’s developed, and the different views of the reality has an impact on the research process.

2.1.1 Ontological considerations

The first paradigm ontology describes the view on nature in our social reality and how its created. Ontology refers to whether we objectively can see the reality, or if we just see the shadows of the world through ideas. Ontological assumptions can be divided in two different categories; constructionism and objectivism (Bryman and Bell, 2015, p. 32). The first view discusses whether social entries can be considered a social construction and if entities can be considered objective by social and external actors. Constructionism regards reality as socially constructed, simply as the result of general actions and human interaction. Social scientists present this reality construction which always can be revised, changed and not seen as definite (Bryman and Bell, 2015, p. 33). Objectivism on the other hand, view the world as if its reality is independent of social actors, and that social phenomena exists whether we research them or not. Objectivism claims that a real reality exists and has a meaning, no matter social actors, as if it’s made of measurable and solid objects, which exists even when we don’t look at them. This point of view regards the research issues as structured and concrete (Bryman and Bell, 2015, p. 32).

This thesis has an objectivistic approach. We are encountered with financial data in our data collection, that is without interpretation possibilities from us as authors, and is existent whether it is gathered for this research or not. We cannot influence original data, and what is to be researched upon has nothing to do with our own intellectual abilities or values. Data is measurable, and statistical process in the method chapter are widely known, mathematical and, without possible interpretations or perceptions. However, when analyzing the results there is a small room for misinterpretations. We as authors are aware of this, therefore are the results carefully concluded upon. We recognize the global warming and rising temperatures, and as a measurable observation in our reality. Hence, the need for sustainable investments and corporate sustainability actions. If we did not recognize climate change, this thesis would have lost its purpose. Contradictory, one could claim that a subjectivist approach would apply, since global warming is a possible result of human interaction, and emissions could be changed upon. This is true, but then that approach would apply to the climate issues, and not this quantitative research method.

There is a need of an ontological discussion simply because views of the reality and ontological assumptions in general affects the development of academic questions and
implementation of research methods. However, in a quantitative study like ours it can be considered less important since our method uses data collection from Thomson Reuters EIKON, and will not deal with any personal data from e.g. conducted interviews. Nevertheless, we still consider it important to regard and to reflect upon these social scientific research components.

2.1.2 Epistemological considerations
The other paradigm epistemology regards theory of knowledge and is divided into two different categories; interpretivism and positivism (Bryman and Bell, 2015, p. 26-29). The difference between them is that the positivism is built on explaining and applying natural science methods in a social context, and interpretivism advocates the usage of subjective meaning of social actors (Bryman and Bell, 2015, p. 28-29). The positivistic epistemology advocates the usage of scientific methods when studying a social reality, with the purpose of providing explanations and identify causes of a phenomenon, to identify the generic in the society. Further, the positivistic view claims that only actual knowledge is possible to observe, and research should be objective, and facts gathering. The purpose with theories is to generate hypotheses which can be empirically tested (Bryman and Bell, 2015, p.28).

The contrasting epistemology interpretivism, are used to understand and interpret human actions and give specific understandings of a social phenomenon in a specific context. Research with this epistemology is bounded to context, time, individual or a group of people and culture. Interpretivism emphasizes the need of a strategy which makes a difference between humans, and research object in science to capture the subjections in social actions (Bryman and Bell, 2015, p. 29-30). Varying research in the same area accumulated, builds new knowledge. In other words, by making general relationships among many different empirically established regularities. Scientific research should work with facts, and the researchers own beliefs, thoughts and values should not affect the quantitative process. The research should therefore be objective. Further, the quantitative research method is characterized as precise with systematic and structured observations, with large set of data collections to give a represented and a maximal reflection of the reality. In our thesis, we use the positivistic epistemology. That is, because the numbers that constitutes financial data and statistical methods cannot be interpreted. Through a systematic literature review on the sustainability area, together with our own interests, we identified a research gap and has formulated hypotheses which can be empirically tested. We work with facts and a large set of data with the purpose to objectively conclude upon our results, through structured observations.

2.2 Research Approach
In our study, a deductive research approach is used. It refers to a classical scientific procedure, that begins from a theoretical framework and goes to finding a research gap, to subsequently formulate hypotheses. These hypotheses can be tested through empirical observations, and then lead to results. The deductive approach is the most common social scientific research approach (Bryman & Bell, 2015, p. 23). The deductive approach states that theories give research. In contrasts, an inductive approach claims that theories are the result of the research. Quantitative research traditionally takes the classical perspective with an objectivistic point of view, along with positivistic epistemology and a deductive approach (Bryman & Bell, 2015, p. 38). Our research and thesis are following this structure. Further, it is important to consider validity and reliability, and to evaluate the quality of those concepts in this thesis. This will be more discussed in chapter eight.
deductive process can be seen as linear according to Bryman & Bell (2015, p. 23). The process is visualized in figure 1.

![Diagram of deductive process](image)

Figure 1. Deductive research method (Bryman et al., 2015, p. 43) modified by the authors.

The quantitative research method claims to give generalizability and precision. However, it also faces critique. Some scientists, for example Schutz (1962), claims that the scientific way of treating social society is inappropriate and gives misleading results. This is because a society can’t be treated with the same methods as when studying phenomena of nature, and one cannot act blind towards those differences (Bryman & Bell, 2015, p. 179).

### 2.3 Perspective

Eurozone companies and their CSR activities will be looked upon with a neutral and external perspective, from an academic point of view. We will investigate the potential relationship between ESG scores and risk-adjusted returns. Thus, we investigate and compare investment strategies based on ESG, such as best-in-class and best-effort approaches. In other words, we examine if the investor could invest in companies that strives to do good and to strives to do better, with concern to ESG aspects, and at the same time possibly gets a financial benefit from investing in these companies. Therefore, we also take a societal perspective, to potentially find academic evidence that could affect and possibly promote corporate sustainability actions and ESG investing.

### 2.4 Choice of theoretical framework

Chosen theories are regarded as relevant for conducting and explaining the idea of this study, and to use as a discussion material with regards to the findings. Theories will be presented in chapter three. The framework consists of theories that could explain the portfolio performance of the strategies in this study. The section of previous literature is regarded as relevant to present how this research subject has arisen and how the research gap was found. Previous research will be presented in chapter four. We have, to our best of knowledge used and presented those theories most appropriate with focus to our thesis. Furthermore, we have also tried to take a stand in the theoretical chapter by arguing why chosen theories and concepts are relevant and how they are being used this thesis.

### 2.5 Systematic literature review

When searching for literature and academic material, scientific databases such as EBSCO and Umeå University library own scientific database were primary used. All literature review in these databases have been conducted by filtering for peer-reviewed articles. The keywords that have been used, among others, are listed below.

*Sustainable investments, CSR, SRI, ESG, sustainable investment strategies, ESG changes, portfolio performance, sustainability ranking, best in class, best effort, risk-adjusted returns, abnormal returns*
Sometimes, as a complement, other databases were used as well, such as Google Scholar, which is a search service that provides scientific journals and publications. All literature used has been critically evaluated to the best of our abilities.

When conducting this thesis, we started with discussing our own interest. Both authors have studied finance at Umeå School of Business and Economics, and both had an interest in sustainability. We started our research in the area of sustainable finance by reading suitable scientific articles, theories and theses to find a research gap that were possible to assess with available and accessible data. We found that a large part of the academic research was conducted on companies, and how their performance was related to their ESG levels. However, not as much were written on changes in ESG and we decided that this should be the subject for our degree project.

2.6 Source Criticism
All sources used in this thesis has been critically evaluated and are considered as relevant. Peer reviewed scientific articles have been a filter in the search for sources, and we have critically reflected upon the sources. Highly cited sources have been prioritized. Further, we have referenced all used material to make the thesis reliable, possible to replicate and to give researchers and authors the credit they deserve. It is very important that the literature used are considered truthful and correct, otherwise it could lead to misleading and inaccurate research. Primary sources are used in first hand, and secondary sources has been used if the primary sources could not be found. The data collection has been made from Thomson Reuters EIKON, and its sub-platform Datastream, which is one of the most comprehensive databases in the world for financial data and corporate information. We have accessed this database through the library at Umeå University. Even though Thomson Reuters EIKON is a secondary source, its considered trustworthy and reliable, and used by many previous social scientific researchers. When evaluating sources, we have looked mainly at them with respect to independence, when they were written, where they were published and their authenticity. Further source criticism of secondary data will be found in the chapter five, due to methodological criticism.
3. THEORETICAL FRAMEWORK
In this chapter, relevant theories for this research are presented, and how they are positioned for later analysis. The first sections cover portfolio theory and market efficiency, which is relevant concepts in studies regarding portfolio performance and abnormal returns. The chapter concludes by describing the theory behind asset-pricing models, which will be used to calculate the risk-adjusted returns in this study.

3.1 Efficient market hypothesis
The idea of market efficiency is one of the most well-known, debated and empirically tested theories in finance. The efficient market hypothesis (EMH) originates from the work of Eugene Fama (1970), who reviewed earlier academic papers and empirical findings concerning financial markets. Fama stated the theory that a capital market is efficient as long as the market prices fully reflect the information available (Fama, 1970, p. 383). Further, Fama divides market efficiency into three different categories; weak form, semi-strong form and strong form. The weak form of market efficiency occurs when the current price reflects all the information set in the historical prices (Fama, 1970, p. 383). This form prevents investors to gain an information advantage based on technical analysis of historical prices, that could result in an advantage and abnormal returns. Further, markets experiences semi-strong efficiency when prices reflect all publically available information (Fama, 1970, p. 383). By this means that whenever new information is publically available, the price will adjust instantaneously to this information. Thus, this form indicates that an investor would not be able to earn any abnormal returns by trading on any public available information. Finally, strong efficiency refers to markets who incorporate all available information in the market prices (Fama, p.383). This includes private information as well, which makes abnormal returns impossible for all types of investors.

The EMH is based on some assumptions that must be fulfilled by the market in order to validate the theory Fama (1970, p. 387). First, there is no transaction costs when trading securities. Second, all information is cost free and fully available to all market participants. Finally, there is a consensus on the market on how to interpret and value the information. Fama agrees to the fact that these assumptions will not be fulfilled completely in practice. However, this is not a necessity for the market to be efficient, as long as enough of the market participants fulfil the assumptions (Fama, 1970, p. 387). In his article, Fama (1970) review empirical evidence and concludes that the efficient market hypothesis in most cases holds true in practice. Ever since the efficient market hypothesis was stated, a vast number of empirical studies have tested how solid the efficient market hypothesis really is. Most research has not been able to support the strong form of market efficiency (Dimson and Mussavian, 1998, p. 96).

This study will investigate the efficient market hypothesis by studying the relationship between levels and changes in ESG scores and risk-adjusted returns. However, a study like this can’t draw any concrete conclusion on whether markets are efficient. This is since any test of the efficient market hypothesis is also a test of the different models used to examine it. If the models are not completely reliable, any results provided by them will also share the same amount of uncertainty. By using public available information in terms of ESG-score and price data, this study will nevertheless focus on indication of the semi-strong form of market efficiency. Supporters of the efficient market hypothesis will argue that it is impossible to earn abnormal returns by implementing ESG-based strategies. As the ESG-score is publically available information, a change in ESG-score and all the
information that potentially comes with it will be reflected in the stock price immediately. Thus, making it impossible to gain any abnormal returns on a strategy based on ESG-levels and ESG-changes. However, if investors misprice information on ESG-scores in the short run, it could argue for the fact that stocks indeed can generate abnormal returns (Renneboog et al. 2008, p. 1734).

3.2 Adaptive market hypothesis

The efficient market hypothesis has as mentioned produced a large body of academic research since it first saw the light of day. Despite this, the question whether financial markets really are efficient remains highly debated and elusive. This is not especially strange, considering the difficulties to measure market efficiency in a complete and reliable way. There has also been a lack of alternative explanations within the field of finance. However, the last decades have seen the rise of behavioral finance, and with it some alternative theoretical explanations regarding the efficiency of financial markets. One of these theories is called the Adaptive Market Hypothesis (AMH), and is developed by Lo (2004).

In the beginning of his article, Lo (2004, p. 17) aims behavioral critique against the EMH. The author points to the fact that psychologist and experimental economist have documented several departures from the cornerstone that EMH rest upon, that is the view of market participants as completely rational and with a constant risk-aversion. As investors experiences uncertainty, it could potentially lead to decision-making that is both irrational and undesirable. Examples of this behavioral biases of investors are overconfidence, overreaction, loss aversion, herding, psychological accounting etc. Because models of efficient markets are based on rational choices, they are also likely to be wrong when market participants act upon this irrational behavior. Lo (2004, p. 19) mean that both psychology and economics in its essence is concerned with human behavior. Nevertheless, the gap between the two fields is surprisingly deep and defined by different characteristics. This has caused problems in the debate regarding EMH, and also in providing alternative explanations that incorporates physiological factors in theories regarding financial markets. Based on these arguments, Lo (2004) saw the need for a new theoretical framework that consider both the school of psychology and economics, which he refers to as the adaptive market hypotheses.

Under the AMH, behavioral aspects of the market participants are considered. In this setting, individual doesn’t always strive or are capable of striving for optimization, which is what neoclassical economic theories calls for. Instead, individuals make choices that are merely satisfactory, rather than optimal (Lo, 2004, p. 22). Also, individuals will also make choices based on past experiences, which is a trial-and-error process that will lead them into reaching optimal solutions. This is an adaptive process that will lead to efficiency over time. However, if the surrounding environment changes, the old solutions may no longer generate optimal outcomes. Thus, a new adaptive process is needed to reach optimal solutions for the new environment. During this process, inefficiencies might appear as changes are adaptive, and not instant. Lo (2004, p. 23) mean that AMH can be seen as a new version of EMH. In this version, prices reflect the information dictated by a combination of environmental conditions and the behavior of the different market participants.

The AMH will according to Lo (2004, p. 24-45) have several concrete implications on financial markets in practice. First, the relation between risk and reward is unlikely to be
stable at all time. As both market population, preferences and institutional aspects change over time, it will change and affect the risk and reward relation through adaptive processes. Second, arbitrage opportunities do exist from time to time. The financial markets existence is dependent upon opportunities to profit, otherwise it would be no point of gathering information and prices would collapse. Third, the success of different investment strategies will be dependent upon the environment. In some environments, they will perform well and in others they will not. This would imply the reappearance of successful investment strategies. This is a contrast to the classical EMH, in where arbitrage opportunities are swiftly and definitely competed away. Fourth, innovation and adaption is the key to survival. As the relation between risk and reward changes, adapting to changing market conditions is the only way to achieve a desired and consistent level of expected returns.

The adaptive market hypothesis might be an alternative hypothesis to EMH in studies regarding the relationship between ESG and risk-adjusted returns. First, the adaptive processes of markets would explain the abnormal returns that ESG-investing has been proven to generate in earlier studies, such as in Kempf & Osthoff (2007) and Statman & Glushkov (2009). This thesis has already touched upon empirical evidence that hints towards investors struggles with pricing ESG. Now, it is possible the market has adapted and learned how to price ESG, which newer studies such as Halbritter & Dorfleitner (2015) indicates. However, it could still be an adaptive process that can provide results that are not aligned with the notion of efficient markets. This is because the concept of ESG and how it should be measured in a complete and satisfying way has not yet received a full consensus by providers and investors. Second, AMH takes behavioral economics into consideration. This can account for the fact that investors can have other objectives in financial decisions than pure economic ones, which will affect the market. For example, the increasing amounts of SRI or ESG-based funds on the market is proof of other market participants that does not only follow the value-maximizing notion of EMH, as they are investing under restrictions not solely based on economic gains. Finally, it also provides justification for reexamining investment strategies again, as the performance of these according to the adaptive market hypothesis could change from one time to another. Based on these insights, the adaptive market hypothesis will be used in this thesis to discuss upon the results. Especially if these results diverge from that of the efficient market hypothesis. That is, the adaptive market hypothesis will be used as an explanatory theory if it is possible to find a relationship between abnormal risk-adjusted returns and a best-effort or best-in-class approach to ESG-investing.

3.3 Modern portfolio theory
The American economist Harry Markowitz is the father of the Modern Portfolio Theory, or mean variance analysis. This is one of the most well-known theories in finance and portfolio management. In his article “Portfolio Selection” from 1952, Markowitz theorizes on how the expected return of a portfolio is maximized for a given level of risk. The recurring theme of modern portfolio theory is the value of diversification. According to the theory, diversification can create portfolios that gives greater expected return for the same amount of risk.

The theory is built on some assumptions about investors as rational value-maximizing individuals. Rational investors view returns as desirable and the variance of the return as less desirable (Markowitz, 1952, p. 77). Based on these assumptions, the investor would choose the portfolio that gives the highest expected return, to the lowest possible risk.
The Markowitz Bullet (Merton, 1972, p. 1856) illustrates how different portfolios generate different expected returns, given different level of risk. The efficient frontier represents the type of portfolios that generates the optimal expected return for a given level of risk. Following modern portfolio theory, the expected return might shrink as the investable universe grows smaller. This is an argument against all portfolio strategies that can be classified as sustainable investments. As pointed out earlier, the most common SRI-strategy is exclusions. Even this strategy shrink the investable universe, as it excludes companies that are not classified as sustainable. This could potentially move the efficient frontier to produce less expected returns for the same level or risk, or vice versa.

Following a best-in-class and best-effort approach, and only invest in top ESG-performers in a universe or in an industry, as well as ESG improvers, follows the same argument to some extent. These strategies will reduce the investable universe through the screening process, and could thereby provide a portfolio with less risk-adjusted returns. The logic behind modern portfolio theory is used by Benlemlih et al. (2018, p. 5465) to argue against a “best-effort” approach. Since this thesis follows a similar approach, modern portfolio theory will be used here with the same intention.

Markowitz (1952, p. 79) logically argues that a portfolio consisting of stocks in the same industry is much less diversified than stocks from different industries. This is a positive aspect of the best-in-class approach, as it includes stocks from the different industries in the investable universe. The best-effort approach is not defined in such matter that it takes diversification into account, but this study will construct the best-effort portfolios based on the portfolio selection process of a best-in-class approach. Through this, the best-effort will be provided with an automatic diversification.

3.4 Asset-pricing models and abnormal returns
To measure whether a strategy based on ESG generates abnormal returns, the actual return need to be compared with the expected return. There are several asset-pricing models to estimate the expected return for portfolios and underlying assets. This section will review some of the most commonly used models and how they are used to calculate risk-adjusted abnormal returns.

3.4.1 Capital asset pricing model (CAPM)
The most known model is the CAPM, which was developed through the work of William Sharpe (1964), John Lintner (1965) and Jan Mossin (1966). The foundation of CAPM builds upon Markowitz (1952) and his modern portfolio theory. CAPM describes the relationship between the expected return and the risk of an asset. The model assumes that investors are looking for the optimal portfolio, with the highest risk adjusted returns. For the CAPM model to properly hold, it builds on the following assumptions regarding investor behavior and market structure (Bodie et al., 2014, p. 304):

1. All investors are rational and mean-variance optimizers. Thus, all investors are following the Markowitz portfolio selection model
2. All investors are planning on holding asset under the same and single period
3. All investors share the same expectations on the assets
4. All investors can borrow and lend money at the risk-free rate, all assets are publicly traded and short positions on all assets are possible
5. All information is known to the public
6. There is no transaction costs or taxes when trading assets
CAPM in its simplicity builds on the relationship between the systematic risk and the expected return. Systematic risk refers to the risk an asset or a portfolio have in comparison to a market portfolio. Systematic risk is measured by beta. If a portfolio has a beta of one, it means that the risk is identical to that of the market portfolio. This risk is therefore also referred to as market risk and is undiversifiable, as compared to unsystematic risk or firm specific risk which can be reduced through diversification. Calculating expected return using the CAPM follows the following formula (Bodie et al., 2014, p. 316):

\[
E(r_i) = r_f + \beta_i[E(r_m) - r_f]
\]

\(E(r_i)\) = Expected return for asset i  
\(r_f\) = Risk-free rate  
\(\beta_i\) = Beta (systematic risk) for asset i  
\(E(r_m)\) = Expected return of the market

Even though the assumptions behind CAPM might seem unlikely in practice to some point, it remains the most commonly used asset-pricing model in finance. Also, there is some early empirical support for the basic notion of the model, such as Friend et al. (1978). Further, Brown and Walter (2013, p. 48-49) points to the fact that companies use CAPM in their capital budgeting and that regulatory agencies use CAPM in pricing. Hence, this makes CAPM relevant in practice and therefore it should not be rejected completely.

3.4.2 Jensen’s alpha
To measure whether a portfolio have produced better returns than expected, the actual return is compared with the estimated return. This is referred to as Jensen’s alpha, from the work of Michael Jensen (1968). Using CAPM to calculate the abnormal return of a single asset or a portfolio of assets can thus be written as (Jensen, 1968, p. 393):

\[
r_i - r_f = \alpha_i + \beta_i[(r_m - r_f)]
\]

\(\alpha_i\) = Alpha, or risk-adjusted performance for asset i  
\(r_i\) = Actual return of asset i  
\(r_f\) = Risk-free rate  
\(\beta_i\) = Beta coefficient for asset i  
\(r_m\) = Actual return of the market

3.4.3 Fama & French three-factor model
Despite the extensive use of CAPM, critique have been aimed at the model throughout the years. Fama and French (1992) pointed to several problems and contradictions with CAPM, backed by with empirical evidence. Fama & French (1992, p. 438) questioned whether the expected return could be estimated by only accounting for the systematic risk. This arguing led them to expand the model into including more firm-specific factors, that better would explain the variation in the returns. They found that firm size and the book-to-market ratio are the most vital factors in forecasting the expected returns (Fama & French, 1992, p. 450). By combining these factors with the original framework of the CAPM, they constructed their three-factor model (Bodie et al., 2014, p. 426-427):
\[
E(r_i) = r_f + \beta_i E[(r_m - r_f)] + s_i E[SMB] + h_i E[HML]
\]

\(E(r_i)\) = Expected return for asset i  
\(r_f\) = Risk-free rate  
\(\beta_i, s_i, h_i\) = Beta coefficients for the different factors  
\(r_m\) = Return on the market  
\(E[SMB]\) = Expected return of small market capitalization stocks minus the expected return of big market capitalization stocks  
\(E[HML]\) = Expected return of high book-to-market stocks minus the expected return of low book-to-market stocks

The SMB factor represents the effect that firm size has on asset prices. It is supposed to capture the excess return that is generated by firms with a lower market capitalization compared to firms with higher market capitalization (Bodie et al., 2014, p.340). The HML factor on the other hand is supposed to capture the excess return of firms with high book-to-market ratio compared to firms with a lower book-to-market ratio (Bodie et al., 2014, p.340). Fama and French (1993, p.54) found the HML factor to be the most predictive factor out of the three in their model.

### 3.4.4 Carhart four-factor model

Carhart’s four-factor model is a further extension of CAPM and Fama-French three-factor model (Carhart, 1997, p. 60-61). The model builds on Jagadeesh (1990) and Jagadeesh and Titman (1993), who presented evidence of the momentum effect. The momentum effect refers to the fact that stock prices tends to move in the same direction over a period. That is, if the stock goes up one month, it tends to continue to move up the next month as well (Jagadeesh, 1990, p. 896). Carhart (1997, p. 79-80) found that the inclusion of a momentum factor better helped to explain the performance of mutual funds. Thus, the extension of the previous models into Carhart four-factor model have the following look (Carhart, 1997, p. 61):

\[
E(r_i) = r_f + \beta_i E[(r_m - r_f)] + s_i E[SMB] + h_i E[HML] + p_i E[WML]
\]

\(p_i\) = Beta coefficient for the WML factor  
\(E[WML]\) = Expected return of past 12 months “winners” minus the expected return of the past 12 months “losers”.

The WML factor represents the average excess returns of firms with good returns over the firms with poor returns over a 12-month period (Carhart, 1997, p. 61). When Jensen (1968) calculated alpha, he used the CAPM and its assumption as core for the formula. As the three- and four-factor models also builds on CAPM, the process of combining the Jensen’s alpha with these models is straightforward. By adding the alpha factor to the models, any abnormal returns can be calculated while controlling for the relevant factors (Carhart, 1997, p. 61). These alphas are being referred to as multi-factor alphas. Thus, the formula for calculating the four-factor alpha which will be used in this study is given by:

\[
r_{it} - r_f = \alpha_i + \beta_i [(r_m - r_f)] + s_i [SMB] + h_i [HML] + p_i [WML]
\]

Derwall et al. (2011) argues that applying the multifactor model is suitable for evaluating the performance. Further, the models are performance measures that are central in many
of the studies of portfolio performance regarding SRI (Derwall et al., 2011, p. 2141). The reason for this thesis to apply the Carhart four-factor model instead of CAPM or the three-factor model is because the authors finds it to be most used in earlier research that are of specific interest to this study. Benlemlih et al. (2018) uses as performance measure for a best-effort approach, as well as Halbritter & Dorfleitner (2015), Kempf & Osthoff (2007) and Statman & Glushkov (2009) for best-in-class approaches.
4. LITERATURE REVIEW

The literature review can be divided in different sections. The first section will provide related academic literature which can argue for the fact that investing based on ESG criteria could provide risk-adjusted abnormal returns. This will be done through a review of academic studies on CSR and the relationship with financial performance, and which value-enhancing capabilities CSR possible can produce. Subsequently, it will cover how investors have been found to react on improvements in CSR in the past. The last section will cover previous studies on ESG performance and risk-adjusted returns based on best-effort and best-in-class approaches.

4.1 CSR, financial performance and value-enhancing capabilities

This literature review will cover the extensive field of academic research on the relationship between corporate social responsibility and financial performance. One body of academic research has focused on the CSR and accounting-based performance, with different approaches to measure accounting-based performance, such as return on assets (ROA), return on equity (ROE) and earnings per share (EPS). Another body of the academic research have instead focused on CSR and market-based performance, such as Tobin’s Q, book-to-market ratio, and various measures of stock returns and portfolio performance. Lo and Sheu (2007) studied whether corporate sustainability has an impact on the market value of large U.S firms between 1999-2002. As proxy for firm value the authors uses Tobin’s Q, which is the ratio of the market value of a firm to the replacement cost of its assets (Lo and Sheu, 2007, p. 351). As proxy for sustainability, the authors are using a dummy variable if the firm are listed on Dow Jones Sustainability Index. Based on regression analysis with several control variables, the authors find that corporate sustainability is significantly associated with a higher market value (Lo and Sheu, 2007, p. 354). In terms of accounting measures, Eccles et al (2008) found empirical evidence that high sustainability companies in U.S outperformed their low sustainability counterparts, measured through both ROA and ROE.

As the literature on CSR and financial performance as mentioned is extensive and very broad, reviews and meta-analysis can be helpful in providing a general picture over the relationship between the two. Orlitzky et al. (2003) provided a meta-analysis of 52 different studies on the relationship between corporate social performance and financial performance. The study covered over 30 years of research on the subject. The major conclusion is that corporate social performance in general seem positively correlated with corporate financial performance (Orlitzky et al. 2003, p. 423). Further findings indicate that social responsibility tended to contribute to the relationship to a greater extent than environmental aspects. Also, the relationship seems to be stronger for accounting-based measures than market-based measures (Orlitzky et al. 2003, p. 419).

Friede et al. (2015) conducts a second-level review over meta-analyses and previous reviews on the relationship between the ESG criteria and corporate financial performance. By doing so, the study is able to combine the results of over 2000 individual studies. The study thereby investigates several measures of corporate financial performance, such as accounting-based performance, market-based performance, operational performance, perceptual performance, growth metrics and different risk measures (Friede et al., 2015, p. 212). The overall findings indicate that ESG criteria and corporate financial performance indeed are positively correlated, at least on average. This general conclusion can be applied on different regions, asset classes and approaches of studies (Friede et al., 2015, p. 225). The strongest findings are related to studies on firm-level. Portfolio-based
studies on the other hand, is considered as an outlier since these show deviating results in comparison to other studies. The authors argue that the mixed findings from portfolio studies can be the reason to why institutional and private investors consider the relationship between ESG criteria and financial performance to be neutral at best. This potential perception by investors could lead to a barrier of a broader uptake of sustainable investing (Friede et al., p. 225).

There is further academic evidence supporting this positive relationship, as academic research also has explored how corporate social responsibility can affect financial performance. Malik (2015) reviews empirical studies of value-enhancing capabilities of CSR. The author conclude that improved CSR can help increasing firm value by promoting employee productivity and thus operating performance. Further, it can have benefits both on capital markets and product markets as well as building a better corporate reputation (Malik, 2015, p. 433).

The academic literature investigating the relationship between CSR and firm performance is as mentioned once again extensive. The literature is also broad, both in terms of measurement of CSR and in terms of the measurements and methods to evaluate financial performance. Due to this, the results from the research is not completely conclusive. Nevertheless, the findings tend to lean towards a positive relationship between the level of corporate social performance and financial performance of firms. There is also evidence on how CSR can influence financial performance. However, there is less evidence on how long time it takes for the firms to reap these benefits. This is also a limitation to this study as it will only cover year to year changes. It is likely that the value-adding capabilities that possibly arises from increasing CSR will be incorporated in the long-run. If these changes will be visible and have an effect in the short-run is difficult to say. This will further be discussed as critique against the study. Nevertheless, there is a lot of academic evidence and theories that supports the fact that improved corporate social responsibility and ESG-score would be value-adding for the firm. At least over time. Based on the efficient market hypothesis though, this would be reflected in the share price immediately as it would be known to the public. However, there are questions to be raised over markets efficiency in pricing corporate social responsibility. The next section of this literature review will be devoted on how investors react upon issues regarding corporate social responsibility.

4.2 Empirical studies on investors reactions to CSR improvements

To find out how investors are pricing improvements in ESG-scores, the best way is to look at previous event-studies on market reactions. There are not many event-studies that directly focuses on changes in ESG-score. At least not that the authors of this thesis have been able to find. However, they are related studies that can be used as a framework to see how investors reacts to CSR-related issues.

Michlik and Rubash (2011) investigated short-term market reactions to 40 North American companies that were either included or delisted from Dow Jones Sustainability Index between 2002-2009. The average market reaction for the included companies were a drop in the stock price of 0.17%. Most of the companies in the study showed a loss in connection to the announcement (Michlik and Rubash, 2011, p. 25). The authors summarize their study by stating that investors do not realize benefits of sustainable strategies immediately. The documented advantages from these strategies are instead realized over a period of years (Michlik and Rubash, 2011, p. 13). However, this is a
study that is only based on 40 companies and stock price observations. It is hard to draw any definite conclusion from this. Kappou and Oikonomou (2014) instead focused on inclusion or deletions from the MSCI KLD 400 Index. This index consists of the 400 companies from the MSCI USA Index that are best adapted to criteria of social responsibility (Kappou and Oikonomou, 2014, p. 540). The final sample consists of 201 inclusions and 77 deletions from the index during the years 1990-2011. The event-window begins 10 days prior to the index decomposition and ends 15 days after it. To measure market reactions, the authors are calculating abnormal returns based on the CAPM. The results show no significant reaction for inclusions during the whole event-window (Kappou and Oikonomou, 2014, p.543). The abnormal returns the day of the inclusion is close to zero. The only significant market reaction is the abnormal return the day a firm faces deletion from the index. This market reaction is quite severe, with negative abnormal returns of -2.02% (Kappou and Oikonomou, 2014, p. 543). Thus, the market show almost no reaction to good CSR-news, while the reaction to bad CSR-news is significantly negative.

The above-mentioned studies are based on U.S firms and only covers data for years up until 2011. Hawn et al (2018) provides a global study based on investor reactions over inclusions or deletions of the Dow Jones Sustainable Index World. The study is based on the years 1999-2015, covers 27 countries and includes 408 inclusions as well as 272 deletions. The authors use a two-day event window to capture eventual news leakage (Hawn et al., 2018, p. 954-955). The findings from this study share similarities with before mentioned studies, but differs in some aspects. The average return for delisted companies is not statistically significant. One the other hand, inclusions in the index is penalized with a significant reduction in the share price of -0.25% (Hawn et al., 2018, p. 958-959). Overall, the findings indicate that investors place little value in inclusions and deletions from sustainable indexes on a global scale as well.

Krüger (2015) takes a different approach than inclusions and deletions from sustainable indexes. Instead he is focusing on more general news regarding CSR for U.S firms between 2001-2007. The data is based on KLD, a ESG-rating agency which collects and provides news regarding CSR. KLD divides articles and corporate information into positive or negative CSR-events, denoted as strengths and concerns (Krüger, 2015, p. 309). These events are categorized into issues regarding community, diversity, employee relations, environment, human rights and products. The final sample consists of 2,116 events, divided into 574 positive and 1,542 negative ones (Krüger, 2015, p.309, p. 311). The findings based on negative CSR-events are the same as for firms being delisted from sustainable indexes. That is, the market reaction for negative CSR-events is a statistically significant reduction in the share price. For a 11-day event window surrounding the negative events, the mean market reaction was -0.88%. As for positive events, the market reaction was quite surprisingly also weakly negative. However, only the 21-day event window was statistically significant, with a minor drop in the mean share price of -0.16% (Krüger, 2015, p.309, p. 315).

Yet another different approach is that of Crifo et al. (2015). They conduct field experiments in where professional private equity investors participate in closed auctions to acquire fictive firms. The purpose of the study was to see whether ESG-disclosure would impact firm valuation and the investment decision. The main findings demonstrated that firms who were looking to improve their corporate social responsibility did not attract the investors any more than other firms. However, firms who failed to
manage ESG-related issues in a notable way were penalized (Cripe et al., 2015, p. 182). Thus, the findings of this different approach points in the same direction as the event-studies discussed above.

The overall conclusions from these studies is that investor doesn’t seem to react much to events regarding improvements in CSR, at least not in the short-run. One argument against these event-studies is the fact that the information could already be known to the market (Michlik and Rubash, 2011, p. 29). Thus, the positive news would then already be incorporated in the share price and the market reaction surrounding the event is therefore insignificant. However, based on the various significant negative reaction on negative events, this argument loses some of its ground. It seems unlikely that most of the positive news would be known beforehand, while the negative news wouldn’t be.

To sum up, investors does not seem to value ESG-improvements straight away. On the other hand, many studies document a positive relationship between ESG and firm performance and value. This would indicate that firms that are improving their ESG-scores are undervalued, and that the benefits from engaging in ESG-related issues would then be incorporated in the stock price over time rather than straight away.

4.3 Previous studies of best-effort and best-in-class approaches
This literature review is of earlier empirical studies regarding best-effort and best-in-class approaches to ESG investing, or at least similar approaches as a best-in-class. There are a lot of other academic work concerning portfolio management and sustainable investing. However, many of these studies uses different criteria in their portfolio selection, such as only exclusions of a small portion of stocks from the universe, or different classifications of what a sustainable firm is. This literature review will only highlight studies which is directly linked to the purpose of this study and is highly cited within this field. Many other portfolio studies also cover fund performance in different settings. These studies are also not considered in the literature review since it can be difficult to draw a complete analyze of the underlying assets in sustainable funds, even though the fund is classified as a SRI-fund. With this in mind, no other studies have to the best of our knowledge been carried out in the Eurozone specifically under the same period as this and with enough of the same characteristics to deem this study unnecessary.

Benlemlih, Jaballah and Peillex (2018)
The study by Benlemlih et al. (2018) is the only academic paper that the authors behind this thesis have found that are focusing on a best-effort approach to ESG investing. In their study, they investigate different financial effects of changes in ESG ratings, or CSR as referred to by the authors. They are basing their data on a sample of 1,621 US firms during the years of 1996-2011. As measure, the authors are using ratings from KLD, which consists of 13 CSR dimensions grouped into seven qualitative issues and six controversial business issues. In their study, the authors focus on the qualitative issues and calculates their own aggregated score. These scores are positive and negative ratings, in terms of strengths and concerns for each of the issues. They are also binary (0/1) for every concern respective strength (Benlemlih et al. p. 5469). This gives a CSR score that ranges between a maximum of seven, and a minimum of seven. The changes utilized in the study are based on the difference between CSR ratings in year t-1 and t-2.

Based on the stakeholder theory, equilibrium models and previous literature, the authors hypotheses that changes in CSR-ratings are negatively related to firm systematic risk and
positively related to firm value (Benlemlih et al. 2018, p. 5468). For portfolio management, the authors argue that a decreased risk following upgrades in CSR scores, enlargement of the investor base and reduced information asymmetry will cause the expected returns of these stocks to decrease (Benlemlih, 2018, p. 5468). Based on this, they hypothesize that a best-effort strategy will generate lower portfolio performance.

To test the hypothesis regarding the changes in CSR and its effect on systematic risk and firm value, they run regressions based on yearly changes in CSR ratings while controlling for different variables. The dependent variable in the regressions is the beta value for systematic risk and book-to-market ratio for firm value. The regressions show that the changes in CSR is statistically significant and negatively related to both firm’s systematic risk and book-to-market value. These findings support both hypotheses stated in the study (Benlemlih, 2018, p. 5475).

To test for portfolio performance, Benlemlih et al (2018, p. 5471) uses two strategies in their portfolio construction. The first strategy consists of a long position in all stocks with an upgrade in ratings and a short position in all stocks with a downgrade. The other strategy consists of a long position in upgraded stocks and a short position in stocks where ratings remain neutral. The portfolios are then rebalanced every year to incorporate the new changes in the ratings. The stocks in the portfolios are equally-weighted. To measure risk-adjusted returns, they use the Carhart four-factor model in their regressions.

The results of the regressions are a negative annual alpha of -3.75% for the first strategy and -4.29% for the second, where both alphas are statistically significant. (Benlemlih, et al., 2018, p. 5479). Further, the authors also test these strategies for all the seven individual CSR issues provided by the KLD ratings. Most of the alphas for the individual issues are negative as well. However, most of them are not statistically significant. The only positive alpha is observed for the issue concerning human resources. The overall conclusions of these findings are in line with the hypothesis stated. That is, a best-effort approach to sustainable investing leads to lower financial performance of portfolios in terms of risk-adjusted returns (Benlemlih et al., 2018, p. 5479).

As mentioned, this study is to the best of our knowledge the only academic paper that considers a best-effort approach. Therefore, the results are of great interest for this thesis. However, there are many differences between the study in this thesis and Benlemlih et al. (2018), besides the different regional context and time period. The first and biggest difference is connected to the portfolio construction and the strategy. Benlemlih et al. (2018) are considering strategies based on long and short positions. This study will only cover a long strategy for the best-effort portfolios. Another difference is that the study in this thesis will consider industries in the strategy, much like a best-in-class approach. Further, they are using a completely different score to utilize CSR. Sources has found that ratings for ESG scores by different providers can differ and influence returns (Halbritter and Dorfleitner, 2015, p. 35). All these differences should be considered when comparing these two studies, as they may provide differences in the result.

**Kempf & Osthoff (2007)**

Kempf & Osthoff (2007, p. 910) used KLD ratings to measure corporate social responsibility and to form portfolios of stocks included in the S&P 500 and DS 400 during 1992-2004. The portfolios are rebalanced every year as new sustainability ratings are being published. They apply the best-in-class approach in their portfolio construction to
overcome bias towards some industries (Kempf & Osthoff, 2007, p. 913). To measure the risk-adjusted performance, they use the Carhart (1997) four-factor model. As for portfolio construction, they create several high-rated portfolios that consist of the top performers in each industry and several low-rated portfolios that consist of the worst performers in each industry (Kempf & Osthoff, 2007, p. 912). The inclusion of companies in each of these portfolios are based on so called cut-off rates. For example, a cut-off rate of 10% mean that the top 10% performers with respect to a specific KLD rating and industry is included in the portfolio. In their study, the authors are using several different cut-off rates. Portfolios are formed within all areas of the KLD ratings as well as for two different combinations of them, and are both equally- weighted and value-weighted.

The main findings of the study show that a best-in-class approach are able to generate significant and positive abnormal returns, with an alpha of up to 8.7% annually (Kempf & Osthoff, 2007, p. 919). This alpha is based on a portfolio with a long position in the top 5% of socially responsible firms and a short position in the bottom 5%, thus a so called long-short strategy. The positive alphas are statistically significant for cut-off rates based on 10% and 25% as well. Since the study conducted in this thesis will only consider a long position strategy, the results from these strategies are of additional interest. Based on a 10% cut-off rate, significant alphas of 2.88% and 2.71% are found for both combinations of the KLD score (Kempf and Osthoff, 2007, p. 917). Positive alphas are also found for several of the individual categories of the KLD ratings. Most of the findings reported in the study are based on value-weighted portfolios. The study in this thesis will only consider equally-weighted portfolios. However, the results appear to be consistent for most of the equally-weighted portfolios as well (Kempf and Osthoff, 2007, p. 918).

The authors further conclude that the best-in-class approach seems to work best when investor restricts themselves to stocks with extreme sustainability ratings. This, since the abnormal returns are higher for the best-in-class approach based on lower cut-off rates in comparison to cut-off rates that includes more stocks in the portfolio selection (Kempf & Osthoff, 2007, p. 921).

**Statman & Glushkov (2009)**

Statman & Glushkov (2009) provides a similar study as Kempf & Osthoff (2007), but with some differences. They are excluding companies that have no indicators of strength and no indicators of concerns, as they argue that these companies are probably not examined by KLD even though they are represented on their list (Statman & Glushkov, 2009, p. 38). Also, they are expanding the time-period until September 2007. In addition to the four-factor model used by Kempf & Osthoff (2007), they are also using CAPM and three-factor model. The findings demonstrated by this study is also similar to the previous one. A portfolio consisting of the top third performers minus the bottom third performers could provide statistically significant abnormal returns on an annual basis, regardless of which performance benchmark being applied. The annual abnormal returns based on CAPM mounted to 3.18%, while the three-factor and four-factor model provided abnormal returns of 6.12% and 5.54% respectively (Statman & Glushkov, 2009, p. 39). The authors connect these findings with the “doing good while doing well” hypothesis, which is stated by the notion that investors and managers underestimate the benefits of being socially responsible (Statman & Glushkov, 2009, p. 34).
Halbritter & Dorfleitner (2015)

Halbritter & Dorfleitner (2015) conducted one of the most comprehensive studies of ESG-investing. They studied the U.S market from 1991 to 2012, while using ESG data from ASSET4, Bloomberg and KLD to be able to compare different outcomes from different providers of ESG score. As for portfolio construction, the authors form long position portfolios consisting of the 20% best performing firms in a category, as well as short position portfolios consisting of the top 20% worst performers (Halbritter & Dorfleitner, 2015, p. 28). In their evaluation, they are considering both long, short and long-short strategies separately. They apply the best-in-class approach by dividing the firms into ten different industries. To evaluate performance, they are in similarity to earlier studies using the Carhart (1997) four-factor model as well as Fama-MacBeth (1973) cross-sectional regression. To further expand and strengthening their study, they are applying different cut-off rates than 20%, using different portfolio weightings as well as evaluating the individual ESG pillars and an overall “best-in-universe” approach (Halbritter & Dorfleitner, 2015, p. 28).

Their study reveals some important findings. First, they mostly find a non-significant relationship between ESG score and abnormal returns. (Halbritter & Dorfleitner, 2015, p. 35). However, some of the portfolios based on certain ratings and cut-off rates provided positive alphas. Second, by dividing their sample into different time periods, the authors show that the outperformance of high ESG-stocks have declined in later time-periods (Halbritter & Dorfleitner, 2015, p. 31). The author argues that this would be the reason to why several earlier studies have documented positive abnormal returns earlier, such as Kempf & Osthoff (2007) and Statman and Glushkov (2009). Last, the influence of ESG-scores on portfolio returns are dependent on which ESG-provider that are being considered (Halbritter & Dorfleitner, 2015, p. 35). Thus, there are differences to be considered when choosing the provider of ESG score.

Van de Velde et al. (2005)

Van de Velde et al (2005) is one of few studies to our knowledge that is investigating the portfolio performance of a best-in-class approach in the Eurozone explicitly. They do so by creating four portfolios based on sustainability ratings provided by Vigeo, an independent agency that screen European companies on CSR (Van de Velde et al. 2005, p. 131). The period of study ranges from the beginning of January 2000 until the end of November 2003. Thus, the study cover a rather short period. For portfolio construction, they create one portfolio consisting of companies in each sector that have a score at least one standard deviation above the mean. They call this the portfolio the “best”. Thus, this portfolio can be viewed as an adaption of the best-in-class approach since they are considering the top performers in different sectors. They also create a portfolio called “good”, which consist of all the companies with scores between the mean and the one standard deviation above. Following this logic, they further create a “bad” and “worst” portfolio that have ratings below the mean and below one standard deviation of the mean respectively. (Van de Velde et al. 2005, p. 132). The portfolios are created for both the overall score and for the individual ratings concerning human resources, corporate governance, environment and society as well as community. As performance measurement, they use the Fama and French three-factor model to assess the risk-adjusted returns of the portfolios.

As for the financial performance of these portfolios, all alphas were found to be insignificant. The “best” portfolio did however produce positive alphas for all the
individual ratings and an alpha of almost zero for the overall rating (Van de Velde et al, 2005, p. 134-135). The authors conclude that the insignificant results of the study may be due to the short time horizon, and that a successful performance from sustainable investing may be proven over a longer period (Van de Velde, 2005, p. 137). The best-in-class approach in this thesis will resemble that of Van de Velde et al (2005) in many of the aspects since it is in the same regional area. It will also differ in some aspects, for example regarding the ESG-ratings and performance measure. By considering a much longer period, this study hopes to provide additional and stronger results of investing according to a best-in-class approach in the Eurozone. Also, to provide evidence of whether a longer period could result in significant alphas, as speculated by Van de Velde (2005, p. 137).

**Auer (2016)**
Auer (2016) focuses on the whole European market during 2004-2012 with ESG data from Sustainalytics. In a critique against earlier studies, the author is arguing for a performance measure based on total risk rather than one based only on market risk. This is because social screens automatically shrink the universe and thereby could expose investors to otherwise diversifiable risk (Auer, 2016, p. 382). Based on this, the study uses the Sharpe ratio to evaluate the performance of the social portfolios. The methodology of portfolio construction is based on various cut-off rates based on a negative screening (Auer, 2016, p. 387). This is applied both for the overall ESG score and the individual components. In a first stage, a portfolio is constructed based only on the criteria that firms are rated. Later, the author form portfolios using negative screens that range from 10% up to 90%. As the exclusion rates increases, the portfolio construction more and more resembles a best-in-universe approach. However, by not focusing on industries, the study does not consider a best-in-class approach (Auer, 2016, p. 393). The findings indicate that social investing is significant better in terms of risk-adjusted returns, but only for low cut-off rates in terms of negative screening. By this definition, a low cut-off rate implies more companies added in the portfolio, i.e. the opposite from the definition used in earlier studies and in this thesis. This finding is eligible both for the combined ESG score as well as the individual score. As for higher cut-off rates, the risk-adjusted performance fall closer toward that of the benchmark (Auer, 2016, p. 395). For the highest cut-off rates, the risk-adjusted performance is the same as the benchmark for the combined ESG score and the governance score. However, the risk-adjusted performance is significant lower than the benchmark for the social and environmental score (Auer, 2016, p. 393).

**4.4 Summary of previous empirical literature**
Table 1 summaries the characteristics and the findings of the reviewed studies concerning best-effort and best-in-class approaches. As stated, Benlemlih et al. (2018) is the only study that investigates a best-effort approach. The findings from that study indicates lower portfolio performance in terms of risk-adjusted returns. Further, the results from a best-in-class approaches is inconsistent over regions and time periods. These empirical findings will be discussed in relation to the findings of this study.
Table 1. Previous empirical literature

<table>
<thead>
<tr>
<th>Authors</th>
<th>Approach</th>
<th>Region</th>
<th>Period</th>
<th>Performance Measure</th>
<th>Risk-adjusted returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van de Velde et al. (2005)</td>
<td>Best-in-class</td>
<td>Eurozone</td>
<td>2000-2003</td>
<td>Fama &amp; French three-factor</td>
<td>Non-significant</td>
</tr>
</tbody>
</table>
5. PRACTICAL METHOD
This chapter will first describe the hypotheses stated to answer the research questions. Subsequently, the methodology leading to these hypotheses being tested will be presented through the sample, data, ESG score, portfolio construction and the performance measure. The section after that covers statistical tools and potential sources of error. The chapter is concluded by a section concerning methodological criticism.

5.1 Hypotheses
This study aims to investigate whether there is a relationship between improvements in ESG score or the absolute level of ESG score and abnormal returns, through a best-effort and best-in-class approach. Statman & Glushkov (2009, p. 34) states three different hypotheses to address the risk-adjusted returns of socially responsible stocks in comparison to conventional stocks or the market. The first they are referring to is the “doing good but not well” hypothesis. The implication of this hypothesis is that the expected returns of socially responsible stocks are lower than the expected returns of conventional stocks. The reason is that the costs of improving social responsibility, and thus the ESG-score of the company, is superior to the benefits. Most investors will know this, and it will be reflected in the returns (Statman & Glushkov, 2009, p. 34). The second is referred to as the “doing good while doing well”-hypothesis. The outcome of this hypothesis is the opposite, that the expected returns of socially responsible stocks exceed those of conventional stocks. For this hypothesis to come true, investors would have to underestimate the potential benefits of being socially responsible. The reason could be that the cost is immediately, but the benefits lies in the future (Statman & Glushkov, 2009, p. 34). Earlier in the study we presented academic evidence that supports the notion that investor does not react heavy to ESG-related issues. The third hypothesis presented by Statman & Glushkov (2009, p. 35) is a “no effect”-hypothesis. Under this hypothesis, the expected returns of socially responsible stocks are the same as the expected returns of conventional stocks. This hypothesis would come true if increasing social responsibility is costless and doesn’t add any value to the company. It would also be true if increased benefits counterbalance the costs of improved social responsibility.

As one can tell, there are argument both for and against abnormal risk-adjusted concerning ESG investing. These hypotheses were based on literature concerning best-in-class approaches. Though, we see arguments applicable on best-effort approaches as well, described earlier in the literature review. In some aspects even more since it concerns the actual change and the actual effort companies do to improve their CSR. Studies based on top performers doesn’t consider at which time the company becomes a top performer, unless it happens during the period researched. The hypotheses that are being tested in this study are as follows. The first hypothesis is regarding the portfolio performance of a best effort approach, and if it has a relationship with abnormal returns. The null hypothesis states that the alpha or the risk-adjusted performance of a best-effort approach is indifferent from zero. The alternative hypothesis states a relation that can be either positive or negative.

\[
H_{01}: \alpha_{BE,ESG} = 0 \\
H_{A1}: \alpha_{BE,ESG} \neq 0
\]

The second hypothesis is similar to the first one, but with regard the portfolio performance of a best-in-class strategy instead.
The third hypothesis concerns the comparison between the two approaches in terms of abnormal returns. That is, whether one of them are significant better in terms of risk-adjusted performance than the other.

\[ H_{03}: \alpha_{BE,ESG} = \alpha_{BIC,ESG} \]
\[ H_{A3}: \alpha_{BE,ESG} \neq \alpha_{BIC,ESG} \]

This study investigates the ESG score as well as the individual pillar score (environmental, social and governance score), to get a full overview of ESG-investing. The same hypotheses are thereby stated for the individual pillar scores as well. These will be tested through portfolios following the same construction as for the overall ESG score, where the only difference is the underlying ESG criteria. See Appendix 1 for full list of hypotheses based on the individual pillar scores.

5.2 Investable universe
As mentioned before, this study aims to investigate sustainable investing, or investing by ESG criteria, in the Eurozone. Most earlier studies have been done on American firms. However, there are as mentioned some studies in Europe. Auer (2016) is using the Europe STOXX 600 as the investable universe. The Europe STOXX 600 is a broad market index, which composes of 600 companies from 17 different European countries. It aims to represent the large, mid, and small market capitalization of the European market. This study will use the Euro STOXX 300 as the investable universe for our European investor. The Euro STOXX 300 is a subset of the STOXX 600 index, and is only represented by the Eurozone companies. The Eurozone countries included in the index are Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain (STOXX, 2019). This will make our Eurozone sample similar to Van de Velde (2005), who investigated approximately 200-300 Eurozone companies between 2000-2003. The population for which this sample refer to is companies in the Eurozone with ESG-coverage by Thomson Reuters. The authors of this thesis have not been able to find the exact amount of companies with ESG coverage in the Eurozone. However, the total coverage in Europe amounts to approximately 1200 companies (Thomson Reuters, 2019, p. 5). Thus, the coverage in the Eurozone is bound to be much less, considering other regions like the United Kingdom constitutes a large part of this coverage. Based on this, the authors of this study argue that the Euro STOXX 300, where most of the companies have ESG coverage, gives a good approximation over the Eurozone. This will be reflected upon further in section 8.3.

For this study, this means that for a firm to be up for portfolio selection the company need to be a member of the Euro STOXX 300 at the time for the yearly rebalancing of the portfolios. Also, the companies need to have a ESG score reported by Thomson Reuters in the year before portfolio selection for the best-in-class portfolios and two years before for the best-effort portfolios, to be able to calculate the change in ESG-score. The Euro STOXX 300 consists of approximately 300 companies. Although, the actual number of companies varies over the years as the index is constantly updated and is a subset of the Europe STOXX 600. Most of these companies have ESG coverage by Thomson Reuters. However, in some cases ESG data is missing for some of the companies, even though the company is included in the index. It is hard to speculate about the reason behind this. It
could be a lack of reporting from the companies, or a lack of coverage from Thomson Reuters at the time being. These firms have simply been ignored for portfolio selection. The amount of missing ESG score is deemed negligible and will likely not affect the results of this study to any large extent. The chosen period for the study is between the beginning of 2009 until the end of 2018, with portfolio rebalance in the beginning of each year. This gives the study a research period of 10 years, and thus 120 monthly return observation for each of the different portfolios.

5.3 Data gathering
The data needed to conduct this study are ESG-scores, stock returns, a proxy for the risk-free rate in the Eurozone as well as data on the factors used in the performance measure of risk-adjusted returns. Several sources have been used to gather this data. This thesis is thereby using secondary data to conduct the study. Constituents of the Euro STOXX 300 along with the leavers and joiners of the index since 2009 have been gathered from Thomson Reuters Eikon. ESG-score, individual pillar scores, country of exchange, Thomson Reuters Business Classification and the company-specific ISIN-code have also been downloaded for these companies through Thomson Reuters Eikon. Data to calculate stock returns have been retrieved through Thomson Reuters Datastream based on the companies ISIN-code. Datastream is a sub-platform to Thomson Reuters Eikon. This study uses monthly stock returns, rather than daily or weekly returns. Monthly stock returns are used in many similar studies earlier, e.g. in Benlemlih et al. (2018), Kempf & Osthoff (2007) and Halbritter & Dorfleitner (2015). As a proxy for the risk-free rate, this study uses the average monthly European Interbank Offered Rate (EURIBOR). Time series for EURIBOR has been downloaded from Deutsche Bundesbank, which is the German Central Bank (Bundesbank, 2019). The EURIBOR is also used by Auer (2014, p. 385) as a proxy for the European risk-free rate. Data used to estimate the factor models has been downloaded from AQR Capital Management (AQR, 2019). AQR provides monthly factor-model data suited for our study for several countries and regions. To be able to construct factors suitable for the Eurozone specifically, data of total market capitalization for respective countries have been gathered from The World Bank database (The World Bank, 2019). This have been complemented with additional data from Datastream to complete the construction of the factors. The procedure of estimating the factor models for the Eurozone will be described more detailed in section 5.7. All the data sources used are being considered as valid and trustworthy sources, as many are used in other research as well.

5.4 Thomson Reuters ESG score
The ESG data collected and used in this study is provided by Thomson Reuters, which is available through their online platform Thomson Reuters Eikon. This data allows the study to be done, as an individual assessment and own collection of data from companies directly would be time consuming and difficult. Thomson Reuters is a leading source of financial news and information. Due to the importance of the ESG score in this study, a review of the dataset and methodology behind it is necessary to provide a complete picture of the study.

Thomson Reuters provides an overall ESG score, as well as scores for the individual pillars of ESG, Environmental, Social and Governance. The company states that the “Thomson Reuters ESG Scores are designed to transparently and objectively measure a company’s relative ESG performance, commitment and effectiveness across 10 main themes based on company-reported data” (Thomson Reuters, 2019, p. 3). These themes
are divided among the individual pillars of the ESG score. The environmental themes refer to innovation, emission and resource usage. The social themes refer to community, workforce, human right factors and product responsibility. The governance themes refer to shareholders, management and CSR strategies. The scores are available both as a letter grading from D- to A+ as well as in a scale of 0-100. This study will use the scale of 0-100 to calculate the numerical change, as it provides a better picture over the actual level and change compared to the grading system.

Thomson Reuters also provides a so called ESG Combined Score (Thomson Reuters, 2019, p. 7). This is a score that discounts for significant ESG controversies that can take place outside of the companies normal ESG activities. If no controversies take place in the year, the ESG score and the ESG combined score are the same. This study will only utilize the ESG score, not the ESG combined score and the controversy score. This can be argued for and against. The combined score may provide a more complete picture over the company in a year, as it includes potential controversies as well. However, this score could disturb our calculations for the best-effort approach in an unsatisfying way. For example, using the combined score in calculating year-to-year changes might provide a change that is the result of a one-time controversy that takes place during the year. The company could then revert to the same level of ESG score in the next year, given that no controversy takes place then. The best-effort approach is intended to capture companies that provide an effort of improving the ESG score on a more permanent basis. Therefore, the choice of the ESG score feels like the more correct one, given the purpose of this approach.

One feature of the scores is that they are benchmarked against Thomson Reuters Business Classifications (TRBC) Industry Group (Thomson Reuters, 2019, p. 3). This is a feature that concerns environmental, social and controversies scores. The governance score is instead benchmarked against the country of the headquarters. This methodology is quite logical, since it is likely that industries will face different challenges and opportunities when it comes to ESG criteria. By this, the raw data takes industry classification into account to some extent. However, since this is a global benchmarking, the use of a best-in-class approach to account for industries would still be necessary when applying the score in a specific region like the Eurozone. Also, the risk exists that the change calculated in this study will be due to changes in the industry benchmark. For example, a company within an industry could receive a positive change in its ESG score due to several other companies within that industry having deteriorated heavily in their ESG-levels. However, since the global ESG coverage by Thomson Reuters is large in terms of companies (over 7000 companies), we assume that this potential effect is unlikely to be too big of a problem when calculating ESG changes for individual companies.

The Thomson Reuters database is continuously updated in alignment with corporate reporting patterns, as new companies are being added for ESG-coverage or following controversy events. This procedure includes recalculations of the ESG score. For most of the companies, ESG reported data is updated once a year in alignment with the company’s own disclosures regarding CSR (Thomson Reuters, 2019, p. 4). This means that ESG-scores could be recalculated historically as well, since all companies are relatively benchmarked. Based on our method and approaches, we find it likely that the companies in this study would be captured even if the historical data is updated slightly. Nevertheless, this could provide some problems concerning an exact replication of this
study in the future. Therefore, the ESG data used in this thesis are available by the authors upon request.

5.5 Data processing and calculations

All the gathered data have been handled in Microsoft Excel, along with any calculations based on the original data. The regressions of the portfolios and the statistical test have been carried out in Stata, which is a software for statistical analysis. Based on the constituent list of Euro STOXX 300 together with the list of leavers and joiners, a list of the exact composition of Euro STOXX 300 at January 1st every year has been created. This procedure guarantees that only companies included in the index at the time for the yearly rebalancing is eligible for portfolio selection. Through this method, the study also takes any problems related to survivorship bias into account, as the list of leavers includes all delisted companies. Survivorship bias refers to the bias in returns of a sample due to excluding past returns on stocks that has left the sample because they were unsuccessful (Bodie et al., 2014, p. 439-440).

Some calculations and adjustments have been made from the original data retrieved. To calculate the stock returns, a monthly return index (RI) for each company were gathered through Datastream. The return index represents an asset's theoretical growth in value over time, with the assumption that dividends are being reinvested continuously. The return index has been used to calculate the monthly returns through the following formula:

\[ R_t = \left( \frac{RI_t}{RI_{t-1}} \right) - 1 \]

\( R_t \) = Return for month t  
\( RI_t \) = Return Index month t  
\( RI_{t-1} \) = Return Index month t-1

As can be noted by the formula, this study uses simple returns. Another option which is common in finance is to use the logarithmic returns, or continuously compounded returns. The reason for why this study uses simple return series is that the factor returns from AQR used as the performance measure for risk-adjusted returns are calculated as simple returns. Also, simple returns have some advantages over continuously compounded returns in aggregation of returns into portfolio returns (Campbell et al. 1997, p. 11). When presenting the results, the monthly alphas of the portfolios are annualized to display the results in a more descriptive manner. The monthly returns are annualized using the following formula:

\[ R_{annual} = (1 + R_{monthly})^{12} - 1 \]

\( R_{annual} \) = Average annualized returns  
\( R_{monthly} \) = Monthly returns

Calculations based on the original data has also been carried out to retrieve ESG-changes for the best-effort approaches, since it is not directly reported by Thomson Reuters. The method follows closely that of Benlemlih et al. (2018, p. 5470), even though they are using another ESG-score than that of Thomson Reuters. Thus, this study defines and calculate ESG changes as:
\[ \Delta ESG_t = ESG_t - ESG_{t-1} \]

\( \Delta ESG_t \) = ESG change in year \( t \)
\( ESG_t \) = ESG Score in year \( t \)
\( ESG_{t-1} \) = ESG score in year \( t-1 \)

The same procedure as stated above follows for calculating the change in the individual pillar scores, i.e. the environmental pillar score, the social pillar score and the governance pillar score.

### 5.6 Portfolio construction

This study aims to investigate whether a best-effort approach to ESG investing can earn abnormal risk-adjusted returns. It also aims to compare this approach to more commonly studied approaches such as best-in-class. To accomplish this, several portfolios have been constructed based on the underlying notion behind these strategies. All the portfolios are rebalanced at January 1st every year and held unchanged until next year. To be able to be included for portfolio selection a company must be included in the Euro STOXX 300 at the time of rebalancing, and to have an ESG-data matching the criteria for the specific portfolio. If a company included in a portfolio is delisted from Euro STOXX 300 during the year it will remain in the portfolio until next year’s rebalancing, given that it is still publicly traded. If the portfolio company instead is delisted from its stock exchange and not publicly traded anymore, the holding is assumed to be sold and reinvested equally over the remaining stocks in the portfolio.

Thomson Reuters ESG disclosures are updated once a year for each company, based on the company’s own ESG disclosures (Thomson Reuters, 2019, p. 4). This means that there is no perfect date for portfolio rebalancing, as ESG-scores for different companies are reported at different times during the year. Therefore, this study chooses to rebalance the portfolio at the beginning of the year, based on the ESG disclosures and ESG-score reported in the year before that. This means that the portfolio selection in the beginning of 2009 is based on the CSR data that is reported by the company and then as an ESG-score by Thomson Reuters sometime during 2008. This reported data is based on the fiscal year 2007 for the companies. In most cases, it is likely that the ESG-data is reported and available for investors earlier than January 1st the year after. However, since there is no specific time requirement for companies to report their CSR-data, this study has chosen this time lag for the portfolio selection. To base portfolio selection to the beginning of the year also resemblance methods used by similar studies, e.g. Benlemlih et al. (2018) and Kempf & Osthoff (2007).

### 5.6.1 Portfolio weighting

All the portfolios in this study are equally-weighted. That is, all the assets have the same weight in the portfolio at the beginning of each rebalancing. This is a common method of portfolio weighting, used in many earlier studies of ESG investing. Benlemlih et al. (2018) uses only equally-weighted portfolios to study a best-effort approach on the US market. Kempf & Osthoff (2007), Statman & Glushkov (2009) and Halbritter & Dorfleitner (2015) all considers equally-weighted portfolios in their studies based on best-in-class strategies. The return of the equally-weighted portfolios in this study have been calculated using the following formula:
$$R_p = \frac{\sum_{i=1}^{n} [R_i]}{n} = \text{Average} \left[ \sum_{i=1}^{n} R_i \right]$$

$R_p$ = Return of the portfolio
$R_i$ = Return of stock $i$
$n$ = Number of stocks in the portfolio

Another common approach used in earlier studies is value-weighted portfolios, which is perhaps more practically common and preferred (Statman & Glushkov, 2009, p 42). To only use equally-weighted portfolio is as mentioned a limitation to this study. Nevertheless, equally-weighting is a common approach alongside value-weighted portfolios in similar studies, as can be seen in the literature review. Further, most investors can make their own choice on how to balance their portfolios in practice. Implications of equally-weighted portfolios will be discussed together with the results.

5.6.2 Cut-off rates
In this study, cut-off rates refer to the percentage of companies included in the portfolio based on the ESG-criteria of the portfolio. For example, a cut-off rate of 10% for a best-in-class portfolio means that the top 10% companies in each class based on the ESG-score is included in the portfolio. Earlier studies have used several different cut-off rates in their portfolio construction. This study will consider 10% and 20% as cut-off rates for the portfolios, since it is commonly used in many earlier studies. Another method could be to define some criteria of what is a high enough ESG change or a high enough ESG score in absolute terms, and to simply include all companies fulfilling these criteria. This would however be somewhat subjective. Also, it could make the number of companies included in the portfolio differ between years and between different portfolios. By using cut-off rates, we ensure that approximately the same amount of companies is included in each portfolio over the period. This makes for easier comparison among different portfolios.

5.6.3 Best-in-class portfolios
The best-in-class approach is based on the idea to invest in the companies performing best in a certain industry or sector. This study uses Thomson Reuters Business Classification (TRBC) to sort companies based on class. TRBC is an industry classification system owned and operated by Thomson Reuters. The system sorts companies into 10 economic sectors, 28 business sectors, 54 industry groups, 136 industries and 837 activities (Thomson Reuters, 2019). This study will consider the 10 economic sectors as different classes in the portfolio selection. The different sectors are Basic Materials, Consumer Cyclicals, Consumer Non-Cyclicals, Energy, Financials, Healthcare, Industrials, Technology, Telecommunication Services and Utilities. This is a similar approach as Kempf & Osthoff (2007) and Halbritter & Dorfleitner (2015), who also uses 10 different industry classes to divide the investable universe based upon industry affiliation.

This study will include the top performers in each class based on the different cut-off rates. Each company in the original sample of Euro STOXX 300 will be considered equal-weighted in this portfolio selection. This means that if there are for example 50 companies in one economic sector at the time for portfolio selection, the five highest scoring companies will be included in the portfolio when applying a 10% cut-off rate. Fractions based on cut-off rates have been rounded to the closest integer. Since the amount of
companies represented in each sector in the Euro STOXX 300 differs, the amount represented in each sector in the portfolio will also differ. This will give the portfolios a closer resemblance to the index and the Eurozone market. See Appendix 2 for a table over average companies included per sector in the portfolios. Another method would be to weight each sector by the market capitalization it has in the index. However, since this study only considers equally-weighted portfolios, this method of inclusion based on equal-weighting of the companies in the index corresponds to that notion.

Based on the above reasoning, one portfolio based on ESG-scores as well as one for each of the individual pillar scores were constructed and rebalanced once a year between 2009-01-01 and 2018-01-01. One portfolio consisting of the top 10% ESG performers in each class, and one based on the top 20% ESG performers in each class. The same applies for the individual pillar scores, i.e. the environmental pillar, the social pillar and the governance pillar. This will amount to a total of eight best-in-class portfolios.

5.6.4 Best-effort portfolios
The best-effort approach is not as easily defined as the best-in-class approach, in which an investor chooses the top performers based on ESG criteria. This study will as mentioned follow Benlemlih et al. (2018) and defines a best-effort approach as the year-to-year change in ESG-scores. However, a best-effort approach could be measured in several different ways. For example, the percentage change in ESG-scores or in absolute terms. By using year-to-year changes in absolute terms, this study treats all ESG-changes as equal. Our approach also doesn’t account for the fact that companies in some cases would see a drop in the ESG-score in the year t-2, and that the positive change is simply the company recovering to earlier levels. Nevertheless, our method of measuring it seems like the most straightforward way. It is also the method that has support in earlier literature as well.

Otherwise, the best-effort approach follows the same portfolio selection as the best-in-class approach, were different economic sectors are considered as classes in the portfolio selection. That is, the top improvers of ESG-scores as well as the individual pillar scores in each class will be included in the portfolios based on the cut-off rates of 10% and 20%. Following this, the best-effort approach also results in eight different portfolios.

5.6.5 Difference portfolios
Another purpose of this study is to compare the best-effort approach to the best-in-class approach. That is, to see whether there is any significant difference in terms of risk-adjusted performance or alpha, as stated by the third hypothesis in this thesis. To answer this, so called difference portfolios of the two approaches has been created. These portfolios are constructed by subtracting the best-in-class portfolio return from the return of the best-effort portfolio counterpart. For example, the return of the best-in-class portfolio based on ESG score and a 10% cut off-rate is subtracted from the best-effort portfolio based on the same criteria. These portfolios are then regressed using the same multi-factor framework as the other portfolios. This method will capture the difference between the alpha of the best-effort approach and the alpha of the best-in-class approach, and provide a p-value of whether the difference in alpha is statistically significant. This method is common in studies of portfolio return evaluation. However, it is often used to compare the investment performance of high-ranked portfolios with low-ranked portfolios when it comes to ESG investing, see for example Derwall et al. (2005). Based
on ESG, the three individual pillar scores and the two cut-off rates, this method will result in eight difference portfolios for comparison between the two approaches.

### 5.6.6 Summary of portfolios

Based on the best-effort approach, best-in-class approach and the difference portfolios constructed to compare the two approaches, a total of 24 portfolios have been created for this study. Table 2 provides a summary over the different portfolios.

<table>
<thead>
<tr>
<th>ESG PORTFOLIOS</th>
<th>GOVERNANCE PORTFOLIOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE ESG (20%)</td>
<td>BE Governance (20%)</td>
</tr>
<tr>
<td>BE ESG (10%)</td>
<td>BE Governance (10%)</td>
</tr>
<tr>
<td>BIC ESG (20%)</td>
<td>BIC Governance (20%)</td>
</tr>
<tr>
<td>BIC (ESG 10%)</td>
<td>BIC Governance (10%)</td>
</tr>
<tr>
<td><strong>ENVIRONMENTAL PORTFOLIOS</strong></td>
<td><strong>DIFFERENCE PORTFOLIOS</strong></td>
</tr>
<tr>
<td>BE Environment (20%)</td>
<td>BE vs. BIC ESG (20%)</td>
</tr>
<tr>
<td>BE Environment (10%)</td>
<td>BE vs. BIC ESG (10%)</td>
</tr>
<tr>
<td>BIC Environment (20%)</td>
<td>BE vs. BIC Environment (20%)</td>
</tr>
<tr>
<td>BIC Environment (10%)</td>
<td>BE vs. BIC Environment (10%)</td>
</tr>
<tr>
<td><strong>SOCIAL PORTFOLIOS</strong></td>
<td><strong>BE vs. BIC Governance (20%)</strong></td>
</tr>
<tr>
<td>BE Social (20%)</td>
<td>BE vs. BIC Social (20%)</td>
</tr>
<tr>
<td>BE Social (10%)</td>
<td>BE vs. BIC Social (10%)</td>
</tr>
<tr>
<td>BIC Social (20%)</td>
<td>BE vs. BIC Governance (10%)</td>
</tr>
<tr>
<td>BIC Social (10%)</td>
<td>BE vs. BIC Governance (10%)</td>
</tr>
</tbody>
</table>

This table presents all the best-effort (BE) and best-in-class (BIC) portfolios in the study. All portfolios are equally-weighted and rebalance 1st Jan each year. 20% and 10% represent the cut-off rate for portfolio selection. The difference portfolios represent the difference in return between best-effort and best-in-class approaches.

### 5.6.7 Transaction costs

Friede et al. (2015, p. 226) links some of the inconclusive results of studies when it comes to sustainable investing to whether transaction costs and fees are accounted for or not. Earlier literature also differs in the aspect of including these costs or not. This study will not assess any transaction costs to the portfolios. One reason not to use transaction costs is since they are likely to differ between different investors, thereby making it hard to retrieve an actual number to generalize transaction costs. Nevertheless, transaction costs exist in the real world of investing and any results should be seen with this under consideration. Therefore, it will be discussed hypothetically when analyzing the results. In the end, it is up to any individual reader or investor to assess whether potential results are still viable when applying transaction costs.

### 5.7 Performance measure

The purpose of this study is to see whether ESG-investing in the Eurozone can lead to abnormal returns or not. The risk-adjusted performance measure used to do this is the Carhart four-factor model. The underlying theoretical notion and origins of the model has been covered in the theoretical chapter. This section cover the practical method of constructing and estimating the factor model. By using the Carhart four-factor, this study follows the method of many other similar studies, e.g. Benlemlih et al. (2018), Kempf & Osthoff (2007), Halbritter and Dorfleitner (2015).
This study will as mentioned earlier use secondary data for the factors in the model, rather than calculating them itself. AQR (Applied Quantitative Research) Capital Management provides these factors through their website (AQR, 2019). AQR is a global investment management firm, with roots and close connection to the academic world. Even though this study uses a set of secondary data, the process of AQR when estimating the different factors need an explanation. The sources and definitions of the process is added in the data file downloaded from AQR. The market factor MKT is the value-weighted return on all the available stocks on a market minus a monthly risk-free rate. The size factor SMB represents the average return on three small portfolios minus the average return on three big portfolios, where the size is measured by total market value of equity. The value factor HML represents the average return on two value portfolios minus the average return on two growth portfolios, where the sorting of stock to either value or growth stocks is measured by companies’ book-to-market ratios. The momentum factor UMD (“up minus down” and earlier referred to as WML) represents the average return on two high return portfolios minus the average return on two low return portfolios, based on the return over the prior 12 months. All portfolios in the calculations of the factors are value-weighted.

AQR provides factor data for Europe as a region, as well as individual factors for many European countries. Since the European factors for the whole region includes many countries that are not a part of the Eurozone, this study has chosen to calculate their own set of Eurozone-factors using the country-specific factors. This means using the individual country factor for Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal and Spain that are all included in the dataset from AQR. To weight these factors by the countries relative size, we collect data of total market capitalization of listed domestic companies for each of these countries from the World Bank. The World Bank is a multi-institution organization that sources funding and knowledge for developing countries (World Bank, 2018, p. 6). The data retrieved from the World Bank is originally from the World Federation of Exchanges database, which is a global industry group for exchanges and clearing houses (WFE, 2019). All retrieved data are denoted in USD. However, the dataset is not complete, since data from Finland and Italy is missing since 2005 and 2015 respectively. This is due to the stock exchanges of these countries leaving the World Federation of Exchanges. To overcome this missing data, a price index denoted in USD for OMX Helsinki and FTSE Italia All Share were downloaded from Datastream. This represent an all-share index for respective countries stock market. By converting the price index to return series, we used this as a proxy for the development of each countries market capitalization. By multiplying the return series to the last value available in the data from the World Bank, we get the complete set of data to compute the weights for each country in the dataset. Each country’s factor-weight in year t is computed by dividing the country’s total market capitalization in year t-1 by the total market capitalization for all included countries in year t-1.

Another adjustment to the data is the need to convert the factors to Euro and add the correct risk-free rate for the Eurozone. AQR has all their factor returns denoted in USD and uses monthly US Treasury Bills as a risk-free rate, for which time series is included in the original dataset. The procedure for converting the factors are conducted as follows. First, the risk-free rate was added back to the MKT factor. Second, return series based on the USD/EUR exchange rate was computed and used to convert the USD returns to Euro for each month using the formula:
EURO Return = (1 + USD Return) * \left(1 + \frac{USD\ Return}{EUR\ Return}\right) - 1

The USD/EUR exchange rate was downloaded from Datastream. Finally, the European risk-free rate was subtracted to reach the markets excess return for the Eurozone. The HML, SMB and UMD factors were converted to Euro returns using the same return formula as above.

Another common source for factor data is Kenneth R. French’s website. However, this website only contains factor data for Europe as a region, not country specific data as this study has used to calculate factors for the Eurozone countries included in the Euro STOXX 300. Another option would have been to calculate the factors by ourselves. This would have been time consuming and unnecessary, as we see no improvements from this option in comparison to using the data from AQR.

5.8 Statistical analysis

5.8.1 Regression analysis

To make a quantitative estimate of the relationship between changes in ESG score as well as levels of ESG scores and risk-adjusted abnormal returns, a multiple time-series regression analysis based on the Carhart four-factor model is conducted. The notion behind multiple regression analysis is whether a response variable is related to two or more explanatory variables (Moore et al., 2011, p. 573). The statistical model used for applying a multiple regression analysis is given by (Moore et al., 2011, p. 591):

\[ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \]

\( y \) = Response variable (or dependent variable)
\( x_1, x_2, x_3 \) = Explanatory variables (or independent variables)
\( \beta_0 \) = Intercept
\( \beta_1, \beta_2, \beta_p \) = Coefficients (or beta)
\( \varepsilon_i \) = Deviation (or error term)

There are some assumptions regarding the error term that should be fulfilled. The error term is assumed to be normally distributed with a mean of 0 and a common, unknown standard deviation that is independent on the values from the explanatory variables (Moore et al., 2010, p. 591). As with most statistical models commonly used in finance, the model is not required to be completely correct. The requirement is that the data do not severely violate the assumptions behind the model (Moore et al., 2010, p. 590).

Using multiple regression analysis in terms of the Carhart four-factor model will translate to the excess portfolio return being the response variable. That is, the variable that the model is trying to explain. The different factor returns are the explanatory variables. Any abnormal return will be represented by the intercept of the regression. Thus, the multiple regression analysis conducted in this study looks like:

\[ r_{pt} - r_{ft} = \alpha_p + \beta_1[MKT_t] + s_i[SMB_t] + h_i[HML_t] + p_i[UMD_t] + \varepsilon_i \]

\( r_{pt} \) = Portfolio return at time t
\( r_{ft} \) = Risk-free rate at time t (EURIBOR 1m)  
\( \alpha_p \) = Risk-adjusted return (alpha) for the portfolio  
\( \beta_t, s_t, h_t, p_t \) = Coefficients of the portfolio on each of the four factors (factor loadings)  
\( MKT_t \) = Return on the market over the risk-free rate (EURIBOR 1m) at time t  
\( SML_t \) = Small minus big, i.e. the return of portfolios with small cap stocks in excess over the return of portfolios with large cap stocks at time t  
\( HML_t \) = High minus low, i.e. the return of portfolios with high book-to-market ratios stocks in excess over the return of portfolios with low book-to-market stocks at time t  
\( UMD_t \) = Up minus down, i.e. the return of portfolios with the past 12 months winners in excess of the return of portfolios with past 12 months losers at time t  
\( \epsilon_t \) = Error term

5.8.2 Hypothesis test
In hypothesis testing, a test of significance will provide evidence of the results against a null hypothesis and in favor for the alternative hypothesis (Moore et al. 2010, p. 367). This method prevents the researcher from ruling out chance as an explanation for the results of the tests. The test is based on a test statistic which is accompanied by a P-value. The P-value is the probability of obtaining a test statistic at least as extreme as the one calculated by the test, given that the null hypothesis is true (Moore et al. 2010, p. 367). Researcher must then choose what level of significance is enough to provide evidence against the null hypothesis. For example, a P-value of 0.07 would considered as statistically significant at a level of 10% but not at a level of 5%.

This study will consider P-values and results based on three levels of significance, which is 1%, 5% and 10%. These levels of significance are common practice when performing statistic tests (Moore et al. 2010, p. 371). They are therefore deemed as important to consider when displaying the results.

5.8.3 Statistical tests
As time-series regressions could be affected by autocorrelation and heteroscedasticity, some statistical test is conducted to check for these issues. All test is carried out in Stata. Autocorrelation refers to the possible correlation between successive values of a time-series, and can cause statistical interference (Moore et al., 2010, p. 708). To check for autocorrelation, this study uses a test developed by Durbin and Watson (1951) called the Durbin-Watson test. The test states the null-hypothesis of no autocorrelation, and provides a d-statistic between 0-4. If the d-statistics is approximately 2, the test indicates no autocorrelation (Studenmund, 2013, p. 334). Additional checks are also conducted by applying the Durbin’s alternative test, which allows regressions that are not completely exogenous. Based on our study, the tests show no strong signs of autocorrelation in the regressions on an overall basis. Even though the tests are based on probability and leaves no guarantees, autocorrelation is thereby assumed to not cause any problems when analyzing the results of the regressions and will not be accounted for further. To check for heteroscedasticity, this study uses a test developed by Breusch and Pagan (1979). Heteroscedasticity refers to when the residuals of a regressions does not have a constant variance. The Breusch-Pagan test states a null hypothesis of a constant variance, i.e. homoscedasticity. It provides a p-value over the null hypothesis based on a chi-squared test. In many of our cases, the p-value is statistically significant and thereby the null hypothesis of constant variance is rejected. Thus, heteroscedasticity can be assumed. To overcome this, the portfolios regressions will be displayed and analyzed using robust standard errors. However, the portfolios have also been regressed without robust standard
errors, and the results are not affected in a way that severely will affect the interpretation of them.

5.8.4 Sources of error
Studies based on hypothesis testing can lead the researcher into making two types of incorrect decisions, referred to as Type I and Type II errors (Moore et al. 2010, p. 382-385. Type I error occurs if the researcher reject the null hypothesis, even though it is true. Thereby accepting the alternative hypothesis. On the contrary, Type II error occurs when the researcher accepts the null hypothesis, even though it is false. Thereby rejecting the alternative hypothesis. These two errors are related, as any attempt to reduce one of them will increase the other. The risk of doing a Type I error can be decreased through a larger sample or by lowering the probability rate. However, lowering the probability rate will automatically increase the risk for a Type II error.

The results in this thesis will as mentioned earlier be displayed with different levels of significance (1%, 5% and 10%). By analyzing the results using all these levels of significance, this study hopes to balance between the risk of Type I and Type II error. Also, the 5% level will be given most weight in the analysis. It is also the most common level of significance out of the three (Moore et al. 2010, p. 371). With these risks in mind, the results always need to be analyzed with caution.

5.9 Methodological criticism
In this chapter, some methodological criticism has already been touched upon. In this section, we are trying to review these and discuss some other sources of critique against the chosen method. The first one is the use of secondary data. Bryman & Bell (2011, p. 320-322) argues for critique against the methodology of using secondary data. One disadvantage is the knowledge about the data set. Since the different data sets in this study is not made by ourselves, the process and authenticity of it cannot be controlled for completely. This critique is not possible to avoid, since it would be nearly impossible to conduct this study not using secondary data within the given timeframe. The sources used for the secondary data is deemed trustworthy, otherwise they wouldn’t have been used. This is confirmed through the extensive use of most of these sources in similar studies and papers.

There is also the risk of human errors when it comes to calculations of different measures that are being used in the study. This risk has been decreased by double-checking the calculations as well as being controlled by both authors. Furthermore, there is the risk of other measures that would reflect what we intend to study in a better way. For example, the changes in ESG scores could be measured in several ways to capture a best-effort approach, as discussed earlier. It is hard to say which measure that is the most appropriate for capturing ESG improvements. We have tried to use measurements that are used in similar studies before, thereby using a measurement like that of Benlemlih et al. (2018).

This study uses only the Carhart four-factor model as a measurement of risk-adjusted returns. Originally, other measurements were also considered to be included, as this could have given the study a more complete picture over risk-adjusted returns. These other measurements were nevertheless not included in a later stage, mainly due to time constraints. Further, the purpose of this study is not to test the Carhart four-factor model and its ability to measure risk-adjusted returns. Therefore, the model and the factors ability to measure risk-adjusted returns under this specific period in the Eurozone is
something that is not specifically assessed in this study. The model is used as a given measurement of risk-adjusted returns due to its supported theoretical notions and its extensive use in earlier financial studies. Due to this reasoning, it is important to highlight that the results could differ if one were to use another measurement of risk-adjusted returns, or if one were to study a different time period or region.

By using a return index from Datastream, this study takes stock splits and dividends into account. Thus, making it in line with practical investing. However, the choice to not include transaction costs and taxes in any of the calculations makes it unpractical in that sense. The reasoning behind this is already covered. Nevertheless, it is a fair point of critique against this study. We are aware that this has an influence on the results, and it will as mentioned be discussed further.

The period in this study cover the years from 2009 to 2018. Under this period, the Eurozone crises has caused financial problems for many of the countries in the region. This is probably something that has influenced the financial performance of some of the companies and possibly the returns and portfolio performance. The full effect of this crisis is hard to draw any conclusions from, but it should be mentioned. Perhaps it is even more important to consider that financial markets overall have witnessed a major upturn in the last decade, in the wake of the global financial crisis. During this period, central banks have exercised expansionary monetary policies that has affected the economy. The Eurozone is no exception of this. It is safe to say that this has influenced the financial markets. It is harder to speculate if and how it influences the strategies and the result of this study. This should nevertheless be taken under consideration regarding the findings, and it is no guarantee that the results would be the same under different market conditions.
6. RESULTS & ANALYSIS

This section will provide descriptive statistics and analysis over the portfolios in terms of ESG scores and cumulative returns. Subsequently, the result of the time-series regressions that estimates the risk-adjusted portfolio performance will be presented. The chapter concludes with a general discussion of the findings.

6.1 Portfolios and ESG score

Table 3 contains a summary over all the different portfolios created. The average score and average change of the score over the whole period is given, with respect to which ESG criteria the portfolio is representing. It can be noted that all portfolios based on a best-effort approach have an average score in the range between 60,1 and 70,7. On the other hand, all portfolios based on a best-in-class approach have score ranging from 79,9 up to 93,1. This is sensible, since the top-performers in each category is unlikely to be the ones that improves the most. This is further confirmed by highlighting the relatively small changes in the score that the best-in-class portfolios have. These portfolios experiences much lower changes in its portfolio-based score in comparison with the best-effort counterpart. Based on this, it seems safe to say that a best-effort approach in the Eurozone includes other companies in the portfolio than the best-in-class approach. Thus, providing an alternative way of ESG-investing that targets different companies.

<table>
<thead>
<tr>
<th>ESG PORTFOLIOS</th>
<th>ESG Score</th>
<th>ESG Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE ESG (20%)</td>
<td>65,07</td>
<td>10,31</td>
</tr>
<tr>
<td>BE ESG (10%)</td>
<td>64,67</td>
<td>12,95</td>
</tr>
<tr>
<td>BIC ESG (20%)</td>
<td>82,00</td>
<td>2,75</td>
</tr>
<tr>
<td>BIC (ESG 10%)</td>
<td>84,73</td>
<td>2,68</td>
</tr>
<tr>
<td>ENVIRONMENTAL PORTFOLIOS</td>
<td>Env. Score</td>
<td>Env. Change</td>
</tr>
<tr>
<td>BE Environment (20%)</td>
<td>70,70</td>
<td>13,43</td>
</tr>
<tr>
<td>BE Environment (10%)</td>
<td>69,43</td>
<td>17,50</td>
</tr>
<tr>
<td>BIC Environment (20%)</td>
<td>90,60</td>
<td>3,32</td>
</tr>
<tr>
<td>BIC Environment (10%)</td>
<td>93,12</td>
<td>3,74</td>
</tr>
<tr>
<td>SOCIAL PORTFOLIOS</td>
<td>Soc. Score</td>
<td>Soc. Change</td>
</tr>
<tr>
<td>BE Social (20%)</td>
<td>68,29</td>
<td>13,52</td>
</tr>
<tr>
<td>BE Social (10%)</td>
<td>67,70</td>
<td>17,34</td>
</tr>
<tr>
<td>BIC Social (20%)</td>
<td>88,40</td>
<td>3,64</td>
</tr>
<tr>
<td>BIC Social (10%)</td>
<td>91,29</td>
<td>4,19</td>
</tr>
<tr>
<td>GOVERNANCE PORTFOLIOS</td>
<td>Gov. Score</td>
<td>Gov. Change</td>
</tr>
<tr>
<td>BE Governance (20%)</td>
<td>60,10</td>
<td>17,08</td>
</tr>
<tr>
<td>BE Governance (10%)</td>
<td>61,97</td>
<td>21,85</td>
</tr>
<tr>
<td>BIC Governance (20%)</td>
<td>79,84</td>
<td>5,46</td>
</tr>
<tr>
<td>BIC Governance (10%)</td>
<td>84,12</td>
<td>7,35</td>
</tr>
</tbody>
</table>

This table presents the score and change in score for all the portfolios over the whole period. The type of score is dependent on which score the portfolio is based upon. The score and changes in score is the average for the companies included in the portfolios.
To put some perspectives over the portfolios score and changes in score, Table 4 provides data over the score and changes for the whole Euro STOXX 300. The table point to average changes that also are much lesser than the ones of the best-effort portfolios. Thus, the best-effort portfolios should capture companies that improves their ESG criteria when making a relative comparison to the Eurozone market or Euro STOXX 300. The score for ESG and the environmental and social pillar are all clear above 50 for the Euro STOXX 300. This is an indication that Eurozone companies on average scores relatively high when it comes to ESG. The score from Thomson Reuters is as stated a relative score, where companies are ranked in comparison to others in that industry group on a global scale.

<table>
<thead>
<tr>
<th>ESG Score</th>
<th>Env. Score</th>
<th>Soc. Score</th>
<th>Gov. Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>62,55</td>
<td>68,58</td>
<td>65,67</td>
<td>52,20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ESG Change</th>
<th>Env. Change</th>
<th>Soc. Change</th>
<th>Gov. Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,60</td>
<td>2,19</td>
<td>1,77</td>
<td>0,73</td>
</tr>
</tbody>
</table>

This table presents score and changes in score for all companies in the Euro STOXX 300, based upon constituents in the index at the beginning of each year. The score and changes in score are averages over the whole period.

6.2 Cumulative returns

Figure 3 displays the cumulative returns for the different ESG portfolios for the entire investment period of 2009-01-01 until 2018-12-31. The cumulative return represents the aggregated return in percentage that an investment in one of the portfolios have yielded over the whole period. The cumulative return of Euro STOXX 300 is also represented in the figure, were capital gains are included through a return index.

As can be seen in Figure 3, the best-effort portfolio based on the top 10% have the highest cumulative return of the ESG portfolios. An investment in this portfolio would have produced a total return of 180,23% over the whole period, which translates to an average annualized return of 10,85%. The other three ESG portfolios performs quite similar. The
best-effort portfolio based on 20% produces a cumulative return of 130.23%, and thus an annualized return of 8.70%. The best-in-class portfolio based on 10% outperforms its bigger counterpart with a cumulative and annualized return of 145.74% and 9.41% respectively, in comparison to 127.64% and 8.57% for the portfolio based on 20%. It can be noted that all portfolios outperformed the Euro STOXX 300 over the period, which earned a cumulative return of 110.12% and an annualized return of 7.71%. One reason behind this could be the fact the equally-weighted portfolios outperform value-weighted portfolios, due to higher weighting of smaller stocks. Based on financial theory, smaller stocks should have a higher risk/return reward. Euro STOXX Index 300 is based on free-float market capitalization (STOXX, 2019, p. 39). Also, based on the purpose of this study it is important to remind that these returns do not account for the risk in the portfolios.

As for the portfolios based on the three individual ESG pillars, the cumulative returns are graphed in figure 4, 5 and 6. For the environmental portfolios, the best-effort portfolio based on 10% performed best by earning cumulative returns of 161.08%. Both the best-effort and the best-in-class portfolio based on a 20% slightly outperformed the Euro STOXX 300 with 128.59% and 120.8% respectively. However, the best-in-class portfolio based on 10% earned only 103.45% in comparison to the 110.12% of the index. This is the only portfolio in the study who did not earn higher cumulative returns than the Euro STOXX 300.

![Environmental Portfolios](image)

**Figure 3. Cumulative returns for the environmental portfolios**
This figure displays the cumulative returns for the environmental portfolios for the whole period.

As can be seen in Figure 5, the social portfolios were all able to outperform the Euro STOXX 300. The best-effort portfolio based on 10% again yielded the highest cumulative returns (173.31%). The best-in-class portfolio based on 10% earned the second highest returns for the period (153.2%). For the portfolios based on 20%, the best-effort portfolio (133.56%) performed slightly better than the best-in-class portfolio (123.82%).
Lastly, the governance portfolios show similar tendencies as the other portfolio, as all portfolios slightly outperforms the market. The best-in-class portfolio based on 10% yields the highest returns (153.79%). However, the returns for the two best-effort portfolios are almost identical, with 150.13% for the one based on 20% and 152.27% for the one based on 10%. The best-in-class portfolio based on 20% earned the lowest returns (132.71%).
BIC Env.10%-portfolio suffers from cumulative returns worse than the index. As discussed, one reason could be the fact that the portfolios are equally-weighted and Euro STOXX 300 are value-weighted. Second, the difference in returns for most of the portfolios is not that excessive. This implies that the chance of many portfolios to earn abnormal risk-adjusted returns seems rather small, unless there is a big difference in the risk of the portfolios. It could also indicate that each of the best-effort and best-in-class approaches by itself contains of a fair amount of the same companies within each pillar of ESG. This would not be a big surprise, since a high ESG score is dependent upon high pillar scores and a big change in ESG is dependent upon a big change in at least one of the pillar scores. It is probably also likely that companies that receives a high ranking in one of the pillar score is also receiving high ranking in the others. Third, the cut-off rate based on 10% is outperforming the cut-off rate based on 20%. The six portfolios producing the highest returns are all based on a cut-off rate of 10%. Last and perhaps most interesting, the best-effort approach is outperforming the best-in-class approach. Based upon the pairing of portfolios as this study uses for creating the difference portfolios, the best-effort approach is producing better returns in seven out of eight cases. However, this should not be a complete surprise either. It is likely that a best-effort approach will target smaller and riskier companies than a best-in-class approach. In classical financial theory, this would indeed implicate higher returns for the portfolios. Next section will in line with the purpose of this study present the results from the regressions based upon the Carhart four-factor model, thereby accounting for the risk in the portfolios.

6.3 Risk-adjusted performance

The results of the time-series regressions based on Carhart four-factor model for the ESG portfolios are reported in Table 5. It is important to remember that the factor models are not based solely upon companies in the Euro STOXX 300, which this study uses as the investable universe for portfolio selection and as benchmark for the cumulative returns. The factors are based upon the stock markets in each country and provided by AQR. Also, the alpha values are annualized for visual purposes. The hypothesis tested by the forthcoming regressions are:

\[ H_0: \alpha_{BE} = 0 \]
\[ H_{A1}: \alpha_{BE} \neq 0 \]
\[ H_0: \alpha_{BIC} = 0 \]
\[ H_{A2}: \alpha_{BIC} \neq 0 \]

As can be seen in Table 5, the four-factor model succeeds in explaining a lot of the portfolio returns, considering the high R^2 ratio of the regressions. All reported R^2 are close to 0,90 or higher. As for abnormal returns, all portfolios report positive alphas. However, only the annual alpha of 3,44% for the smaller best-effort portfolio is weakly significant. This means that the null hypothesis for this portfolio can be rejected using a 10% significance level. For the other portfolios, the null hypothesis cannot be rejected. The market risk is as expected significant and relatively close to one for all portfolios. The beta exceeds one for all portfolios though. Thus, all portfolios have a risk exposure that slightly riskier compared to the market. The coefficient on SMB is highly significant for both best-effort portfolios. This could partly be explained by the portfolios being equally-
weighted, where smaller companies receive a larger portion in the portfolios. However, it also implies a bias in these portfolios towards smaller companies compared to the best-in-class portfolios, since the coefficient for these portfolios are much smaller and insignificant. Also, the UMD coefficient is highly significant for all portfolios, indicating a bias towards stocks with poorer performance in the last 12 months.

Table 5. Time-series regressions for ESG portfolios

<table>
<thead>
<tr>
<th>PORTFOLIO</th>
<th>Alpha</th>
<th>MKT</th>
<th>SMB</th>
<th>HML</th>
<th>UMD</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE ESG 20%</td>
<td>0.0174</td>
<td>1.047***</td>
<td>0.176***</td>
<td>-0.034</td>
<td>-0.113***</td>
<td>0.931</td>
</tr>
<tr>
<td></td>
<td>(0.246)</td>
<td>(0.000)</td>
<td>(0.003)</td>
<td>(0.535)</td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td>BE ESG 10%</td>
<td>0.0344*</td>
<td>1.066***</td>
<td>0.301***</td>
<td>-0.109</td>
<td>-0.108***</td>
<td>0.894</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.101)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>BIC ESG 20%</td>
<td>0.0141</td>
<td>1.083***</td>
<td>-0.037</td>
<td>0.185***</td>
<td>-0.098***</td>
<td>0.971</td>
</tr>
<tr>
<td></td>
<td>(0.186)</td>
<td>(0.000)</td>
<td>(0.334)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td>BIC ESG 10%</td>
<td>0.0182</td>
<td>1.102***</td>
<td>0.023</td>
<td>0.064</td>
<td>-0.069***</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td>(0.141)</td>
<td>(0.000)</td>
<td>(0.568)</td>
<td>(0.214)</td>
<td>(0.002)</td>
<td></td>
</tr>
</tbody>
</table>

This table presents the results of the Carhart (1997) four-factor model. Alpha-values are annualized and based on monthly returns. P-values are reported in brackets and are estimated using robust standard errors. ***, ** and * indicate a significance level of 1%, 5% and 10%.

As for the portfolios based on the individual pillar scores, the results from the Carhart-four-factor regressions are reported in Table 6. The results resemblance that of the ESG portfolios in many aspects. The four-factor model continues to have a high explanatory power of the portfolio returns, viewed through the high R² ratios of the different portfolios. All portfolios report positive alphas except one. The exception is the smaller best-in-class portfolio for the environmental pillar score. However, the negative alpha is very close to zero and has a very high p-value of 0.75. As in comparison to the ESG portfolios, most alpha values are also insignificant for the individual pillar portfolios. This means that the null hypothesis for most of the portfolios cannot be rejected. However, some portfolios report significant alphas. Both the social and the environmental best-effort portfolios based on a 10% cut-off rate have significant alphas of 3.40% and 3.95% respectively. Both portfolios have p-values below 0.05. Thus, the null hypothesis can be rejected for these two portfolios using a stronger level of significance of 5%. The results for the governance portfolios are the ones that diverges most from that of the others. All portfolios report weakly significant alphas, except the smaller of the best-effort portfolios. The null hypothesis for the other three can be rejected based on a level of significance of 10%. This stand in contrast with the other portfolios, where the smaller best-effort portfolios are the only one that is significant. In interpreting the result for the two best-effort portfolios based on the governance pillar, one need to be careful. First, the p-values for the two portfolios are just below and above the significance level of 10%. Second, even though insignificant, the alpha of the smaller best-effort approach is still higher than for the bigger portfolio with 3.3% compared to 2.76%.

As for the different factors loadings, the results of the regressions also resemblance that of the ESG portfolios. Almost all portfolios have a market risk that exceeds one, even though it is close to one in most cases. All best-effort portfolios have a positive and significant coefficient in respect to the SMB factor. This implies a bias towards small capitalization stocks in these portfolios. As mentioned before, a part of this could originate from smaller companies receiving larger weights due to the method of equally-
weighted portfolios. However, this doesn’t seem to be the whole explanation for these portfolios as well since most of the best-in-class portfolios does not have a significant exposure to the SMB factor. Most of the best-in-class portfolios instead have a positive and significant coefficient in respect to the HML factor. This would imply an exposure towards companies with higher book-to-market ratios. The UMD factor is also significant for all the individual pillar portfolios, indicating an exposure towards companies with poorer performance in the past twelve months as well.

Table 6. Time-series regressions for individual pillar score portfolios

<table>
<thead>
<tr>
<th>PORTFOLIO</th>
<th>Alpha</th>
<th>MKT</th>
<th>SMB</th>
<th>HML</th>
<th>UMD</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE Env. (20%)</td>
<td>0,0123</td>
<td>1,106***</td>
<td>0,242***</td>
<td>-0,087</td>
<td>-0,106***</td>
<td>0,941</td>
</tr>
<tr>
<td>(0,375)</td>
<td>(0,000)</td>
<td>(0,000)</td>
<td>(0,134)</td>
<td>(0,000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE Env. (10%)</td>
<td>0,0340**</td>
<td>1,071***</td>
<td>0,278***</td>
<td>-0,095</td>
<td>-0,159***</td>
<td>0,919</td>
</tr>
<tr>
<td>(0,040)</td>
<td>(0,000)</td>
<td>(0,000)</td>
<td>(0,155)</td>
<td>(0,000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC Env. (20%)</td>
<td>0,0135</td>
<td>1,103***</td>
<td>0,067</td>
<td>0,099**</td>
<td>-0,132**</td>
<td>0,961</td>
</tr>
<tr>
<td>(0,275)</td>
<td>(0,000)</td>
<td>(0,134)</td>
<td>(0,024)</td>
<td>(0,000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC Env. (10%)</td>
<td>-0,0048</td>
<td>1,143***</td>
<td>0,035</td>
<td>0,081</td>
<td>-0,056***</td>
<td>0,939</td>
</tr>
<tr>
<td>(0,750)</td>
<td>(0,000)</td>
<td>(0,501)</td>
<td>(0,142)</td>
<td>(0,022)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE Soc. (20%)</td>
<td>0,0198</td>
<td>1,078***</td>
<td>0,194***</td>
<td>-0,005</td>
<td>-0,140***</td>
<td>0,946</td>
</tr>
<tr>
<td>(0,144)</td>
<td>(0,000)</td>
<td>(0,000)</td>
<td>(0,916)</td>
<td>(0,000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE Soc. (10%)</td>
<td>0,0395**</td>
<td>1,050***</td>
<td>0,229***</td>
<td>-0,053</td>
<td>-0,154***</td>
<td>0,919</td>
</tr>
<tr>
<td>(0,023)</td>
<td>(0,000)</td>
<td>(0,000)</td>
<td>(0,351)</td>
<td>(0,000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC Soc. (20%)</td>
<td>0,0116</td>
<td>1,083***</td>
<td>-0,028</td>
<td>0,183***</td>
<td>-0,095**</td>
<td>0,972</td>
</tr>
<tr>
<td>(0,266)</td>
<td>(0,000)</td>
<td>(0,388)</td>
<td>(0,000)</td>
<td>(0,000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC Soc. (10%)</td>
<td>0,0230</td>
<td>1,081***</td>
<td>0,041</td>
<td>0,113**</td>
<td>-0,089**</td>
<td>0,944</td>
</tr>
<tr>
<td>(0,130)</td>
<td>(0,000)</td>
<td>(0,332)</td>
<td>(0,022)</td>
<td>(0,021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE Gov. (20%)</td>
<td>0,0276*</td>
<td>1,049***</td>
<td>0,207***</td>
<td>-0,003</td>
<td>-0,140***</td>
<td>0,928</td>
</tr>
<tr>
<td>(0,090)</td>
<td>(0,000)</td>
<td>(0,001)</td>
<td>(0,956)</td>
<td>(0,001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE Gov. (10%)</td>
<td>0,0330</td>
<td>0,990***</td>
<td>0,164**</td>
<td>0,015</td>
<td>-0,142**</td>
<td>0,874</td>
</tr>
<tr>
<td>(0,115)</td>
<td>(0,000)</td>
<td>(0,025)</td>
<td>(0,838)</td>
<td>(0,002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC Gov. (20%)</td>
<td>0,0248*</td>
<td>1,099***</td>
<td>0,109**</td>
<td>0,116**</td>
<td>-0,186**</td>
<td>0,958</td>
</tr>
<tr>
<td>(0,074)</td>
<td>(0,000)</td>
<td>(0,020)</td>
<td>(0,020)</td>
<td>(0,000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC Gov. (10%)</td>
<td>0,0328*</td>
<td>1,118***</td>
<td>0,118**</td>
<td>0,062</td>
<td>-0,173**</td>
<td>0,936</td>
</tr>
<tr>
<td>(0,073)</td>
<td>(0,000)</td>
<td>(0,033)</td>
<td>(0,318)</td>
<td>(0,000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table presents the results of the Carhart (1997) four-factor model. Alpha-values are annualized and based on monthly returns. P-values are reported in brackets and are estimated using robust standard errors. ***, ** and * indicate a significance level of 1%, 5% and 10%.

The results of the time-series regressions of the difference portfolios are displayed in Table 7. This is conducted with the purpose of comparing the best-effort approach to the best-in-class approach in terms of generating risk-adjusted returns. The hypothesis stated for this is:

\[ H_{03}: \alpha_{BE} = \alpha_{BIC} \]
\[ H_{A3}: \alpha_{BE} \neq \alpha_{BIC} \]
All alpha values except one is insignificant, and with large p-values. The difference portfolio based on the two smaller environmental portfolios is the only exception. The alpha of this portfolio is positive (3.90%) and significant. Thus, the null hypothesis for this portfolio would be rejected based on a level of confidence of 5%. All other null hypothesis would not be rejected based on any level of the significance considered in this study. See Appendix 3 for full regressions of the difference portfolios.

### Table 7. Time-series regressions for difference portfolios

<table>
<thead>
<tr>
<th>PORTFOLIO</th>
<th>Alpha</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE vs. BIC ESG (20%)</td>
<td>0.0033</td>
<td>0.817</td>
</tr>
<tr>
<td>BE vs. BIC ESG (10%)</td>
<td>0.0159</td>
<td>0.439</td>
</tr>
<tr>
<td>BE vs. BIC Environment (20%)</td>
<td>-0.0012</td>
<td>0.930</td>
</tr>
<tr>
<td>BE vs. BIC Environment (10%)</td>
<td>0.0390**</td>
<td>0.047</td>
</tr>
<tr>
<td>BE vs. BIC Social (20%)</td>
<td>0.0082</td>
<td>0.538</td>
</tr>
<tr>
<td>BE vs. BIC Social (10%)</td>
<td>0.0162</td>
<td>0.363</td>
</tr>
<tr>
<td>BE vs. BIC Governance (20%)</td>
<td>0.0027</td>
<td>0.867</td>
</tr>
<tr>
<td>BE vs. BIC Governance (10%)</td>
<td>0.0002</td>
<td>0.994</td>
</tr>
</tbody>
</table>

This table presents the results of time-series regression on the difference portfolios. The alpha-value is annualized and represents the difference in alpha between the compared portfolios. P-values are reported and estimated using robust standard errors. ***, ** and * indicate a significance level of 1%, 5% and 10%.

### 6.4 General discussion

The only other study that the authors behind this thesis has found that are evaluating the risk-adjusted performance of a best-effort approach is Benlemlih et al. (2018). The results from that study differs in comparison to the ones from this, as does the methodology. Benlemlih et al (2018, p. 5478) finds a negative alpha for almost all best-effort portfolios. This study finds a positive alpha for all the portfolios, where four of them are significant when applying a confidence level of 10%, and two of them when applying a confidence level of 5%. However, Benlemlih is studying strategies based on long-short portfolios, that is the abnormal return of KLD upgraded stocks in comparison with KLD downgraded stocks. This study only considers long portfolios based on ESG changes and cut-off rates in excess over the risk-free rate. Other characteristics of each study makes a complete comparison difficult as well, since Benlemlih et al. (2018) base their sample on US firms between 1996 and 2011. Also, they are using a different ESG, or CSR, measure in KLD scores. Further, their portfolios include all companies that received an upgrade in the score, without any stated industry affiliation to consider. Even though these aspects distinguish the two studies, one overall comparison can be drawn. The conclusions of Benlemlih et al. (2018, p. 5479) is that a best-effort approach seems to lower the financial performance of portfolios, at least in the U.S. The same cannot be said as for the financial performance of our best-effort portfolios in the Eurozone. None of the eight portfolios experience a negative alpha. Four of them have a positive alpha which is significantly different from zero based on a confidence level of 5% and 10%. Three out of these four portfolios are based on a cut-off rate of 10%, which include both portfolios that are significant with a level of significance of 5%. Thus, applying a lower cut-off rate seems like a better option to increase portfolio performance than applying higher cut-off rates. This notion that lower cut-off rates produces better returns is aligned with the findings of Kempf & Osthoff (2007, p. 917) for best-in-class approaches. They found that lower cut-off rates produced better returns than higher cut-off rates, thereby urging investors to...
concentrate on the very best stocks in respect to level of ESG. This study would argue the
same for best-effort approaches in the Eurozone.

As for the best-in-class portfolios, this study observes insignificant alphas except in two
cases. The two governance portfolios are the exceptions. The alphas of these portfolios
are positive and significant, but only based significance level of 10%. These findings are
to a high degree in line with Halbritter and Dorfleitner (2015, p. 30), in which abnormal
returns for best-in-class portfolios are found to be insignificant in a clear majority of ESG
portfolios, as well as for different cut-off rates. Considering the Eurozone, these
insignificant findings are also in line with those of Van de Velde (2005), who documented
positive but insignificant alphas for most the portfolios. Van de Velde (2005, p. 137)
speculated that a study covering a longer period would perhaps be able to find significant
alphas, since the successful performance of best-in-class portfolios are more long-term
oriented. This study cannot confirm this speculation. The findings of the best-in-class
approach is thereby in line with the efficient market hypothesis at first sight. However,
both Kempf & Osthoff (2007) and Statman and Glushkov (2009) have earlier documented
positive alphas for best-in-class methods. This could be argued as support for the adaptive
market hypothesis, rather than the efficient market hypothesis. Earlier studies have
documented positive and significant risk-adjusted returns for best-in-class portfolios,
while studies covering more recent time-period does not. Investor perhaps needed the
adaptive process described in the adaptive market hypothesis to be able to price different
levels of ESG correctly. Further, the significant abnormal returns for some of the best-
effort portfolios could be an argument for the fact that this adaptive process is not fully
completed.

One should also remember the market conditions that stock markets have experienced
since the financial crisis, with heavy monetary policies and steep upturns in stock prices.
Perhaps a best-in-class approach is better suited for other market conditions, for example
under the market conditions when earlier studies have found significant abnormal returns
for this approach. On the opposite, it could be that a best-effort approaches instead
produces better results in market conditions like the ones experienced in this study. This
could explain the negative result for a best-effort approach in Benlemlih et al. (2018),
where most of the study concerned the period preceding the financial crisis. This arguing
would also be in line with the notion of the adaptive market hypothesis. Thus, the adaptive
market hypothesis implicates that investment strategies can experience different degrees
of success under different financial conditions (Lo, 2004, p. 24-25). Therefore, it is
important to restate that there is no guarantee that the results of this study would be the
same under other market conditions.

It is important to point to the fact that the method of the four-factor model controls for
the risk factors concerning market, size, value and momentum. Some general
observations regarding these risk factors can be seen based on the regressions. First, all
portfolios except one have a market risk bigger than that of the market portfolio. This is
probably to some extent explained by the equally-weighted portfolios. Thus, smaller and
riskier companies are receiving larger weights than in the market factor, which is value-
weighted. This could explain why most of the portfolio produce better cumulative returns
than the market, with larger weights for smaller companies with a higher risk/return
reward. Second, all portfolios have a significant negative exposure toward stocks with
worse performance in the last 12 months. The reason behind this is harder to draw any
general conclusions from. It should also implicate lower actual returns of all portfolios,
which is not the case. The cumulative return based on the UMD factor over the whole period implicate that a momentum effect is present. We can only assume that the higher market risk of the portfolios, as well as the exposure towards the other two factors are counterbalancing this in terms of the actual returns. The Carhart four-factor are nevertheless controlling for this risk when it comes to the risk-adjusted returns.

An argument against sustainable investing overall is that the screening based on ESG criteria automatically prevent investors from choosing the optimal portfolio. This argument comes from the notion of modern portfolio theory, where all rational investors should hold the market portfolio. Benlemlih et al. (2018, p. 5465) argues that this argument could be particularly important when considering best-effort portfolios, since the amount of stocks with change in ratings are fewer than stocks where ratings are stable. This could lead to the risk of less diversified portfolios. However, our approach is considering diversification by combining the selection process from a best-in-class approach with a best-effort approach. Even through this selection process, the average change in the portfolio score is above ten for all portfolios. This change is above the change required for companies to receive a rating upgrade by Thomson Reuters, if the grading letters are being considered instead. Based on this, the argument of lower diversification based on modern portfolio theory is in many aspects considered by both approaches, without losing the different purpose. This argumentation is strengthening by the results, as no approach seem to lower portfolio performance.

The difference portfolios provide insignificant results in seven out of eight cases. The notion behind this were to see whether a best-effort or best-in-class approach would be superior to the other in terms of risk-adjusted performance. Based on the results, this argument is not supported, except for the environmental portfolios with a lower cut-off rate. The insignificant result does not come as a surprise, considering that most portfolios provided positive alphas in the analysis of them based on earlier regressions. The significance of the environmental difference portfolio can be derived partly from the underperformance of the best-in-class portfolio. This was as stated the only portfolio with a negative alpha of all the portfolios in the study.

One important aspect need to be discussed with consideration of the results in general, and perhaps for the best effort approach in particular. That is the likely effect of introducing transaction costs. As explained before, this study ignores transaction costs, such as fees and taxes. The reasons have already been covered in earlier sections. Nevertheless, a hypothetical discussion of transaction costs is important to provide to readers and potential investors. It can be assumed that a best-effort approach is much more likely to induce larger transaction costs than a best-in-class approach. This is because the turnover rate due to portfolio adjustments are prone to be much larger for this approach. Top performing companies in ESG or individual pillar scores could very well remain top performers for a long period, as ESG scores for companies are proven to be quite stable over time. Thus, the portfolio selection for this approach could be assumed to be quite stable over time. The same argument cannot be said for a best-effort approach. It is not likely that improvers within ESG scores or individual pillar scores are able to keep improving over a longer period of time. Kempf and Osthoff (2007, p. 916) found the turnover rate to be around 34% for their best-in-class portfolios, and that investors could still achieve abnormal returns after adjusting for transaction costs. What the exact turnover rate due to portfolio adjustments would be for our best-effort portfolios is difficult to say. However, by briefly examining the portfolios, we can strongly assume
that it would be higher than 34%. This leads us to questioning whether the alphas for these portfolios would still be positive after applying a realistic amount of transaction costs. Remember, this is only a hypothetical discussion on the effect of transaction costs. Therefore, we would urge eventual future studies to include some sort of approximation of transaction costs in their estimations.
7. CONCLUSION
This chapter provides a conclusion of the empirical study based on the research questions of this thesis. Subsequently, theoretical and practical contribution will be considered followed by a suggestion for future research. The chapter ends with ethical and societal aspects of the research.

7.1 Concluding remarks
Sustainability as a concept is something that has received much more attention in the last decades, and with it the concept of sustainable investments. Both in practice as well as in academic research. However, the opinion of whether it is possible to combine sustainable targets with financial targets is still open for debate. In practice, the value of assets under management for sustainable investment strategies is continuously increasing (Eurosif, 2018, p. 22). Two of these strategies are the best-in-class and best-effort approach. The best-in-class approach to sustainable investing has been tested empirically in many earlier articles. However, the results have been somewhat inconclusive. The best-effort approach is a strategy that have received far less attention. The purpose of this thesis was to see how these strategies affects the risk-adjusted portfolio performance. To fulfill the purpose, three different research questions were stated and the Eurozone was chosen as the region to conduct the study in.

The first research question was “is there a relationship between abnormal returns and a best-effort approach to ESG-investing in the Eurozone?” Based on our approach and the model used for assessing risk, the answer is that a relationship partially exists. However, this is mostly true when only companies with the biggest improvements are being considered. The relationship seems strongest for the environmental and social factor. As for the governance factor, we are careful with interpreting the result to whether a relationship exists, as only the bigger portfolio shows weakly significant results.

The second research question was “is there a relationship between abnormal returns and a best-in-class approach to ESG-investing in the Eurozone?”. Here, our results are in line with more recent findings of best-in-class approaches. That is, there is mostly a non-significant relationship. The governance factor is the exception with a weakly significant alpha for both portfolios. This could be an indication that the governance factor is of more importance to consider for best-in-class strategies.

The third research question was based upon a comparison between the two approaches and was stated as “is there a difference between a best-effort approach and a best-in-class approach in terms of abnormal returns in the Eurozone?”. The answer here is that the approaches in most cases are indistinguishable when it comes to the relationship with abnormal returns and portfolio performance. We do find evidence that a best-effort approach is significantly better than a best-in-class approach for the environmental factor of ESG. Though, this is only true when the top companies in each of the two environmental approaches is considered.

Even though most of our portfolios display insignificant results, the alphas of all portfolios except one are above zero. More importantly, none of them provide significant negative alphas. This gives insight to whether or not investors have to sacrifice returns in order to invest in a more sustainable way. The answer to this seems to be no. This is at least good news for sustainable investing on an overall basis.
7.2 Theoretical and practical contribution
This study has several theoretical contributions. It has expanded the empirical evidence of the relationship between ESG investing and risk-adjusted performance. Foremost, it has given new insight to the implication of a best-effort approach on portfolio performance based on risk-adjusted returns. The results of the empirical evidence are that a best-effort approach does not lower portfolio performance. Instead, the results lean towards that a best-effort approach in some cases can improve the portfolio performance. It also supports the empirical evidence that a best-in-class method to ESG investing in most cases does not produce abnormal risk-adjusted returns.

These findings have practical implications as well. It provides justification for a best-effort approach to ESG-investing. Based on the discussion on transaction costs and the effect that these are assumed to have on portfolio performance of the best-effort approaches in practice, this study does not recommend an exact replication of the portfolios in this study. However, the results favor a best-effort approach as good complement to investors that want to construct portfolios with ESG criteria under consideration. It also provides justification for investing based on a best-in-class approach, since the results does not indicate lower portfolio performance in terms of risk-adjusted returns.

7.3 Suggestion for future research
During this thesis, several ideas for future studies has crossed the mind of the authors. The ideas are mostly connected to best-effort approaches to ESG investing. Some points that would improve this study has already been touched upon in this thesis. This study did neither consider value-weighted portfolios nor transaction costs, at least not more than in a hypothetical discussion. These two factors would have made the study more practically oriented than it is. Therefore, we would once again urge future studies to consider this in their studies. Also, there are other ways to measure risk-adjusted returns that would be interesting to consider. This study is narrow in that sense, since it only considers one approach in the Carhart four-factor model.

During this study, we also considered best-effort approaches based on other aspects than year-to-year changes. It would be interesting to study best-effort approaches that captures companies with longer improving trends in their ESG criteria. This could for example be related to best-effort approaches that captures the best improvers over a period of years. Another idea is to investigate if there are any differences in improving ESG based on whether the initial level of ESG is high or low. On an overall basis, we would like to see more studies over other regions and other periods that concern best-effort approaches.

7.4 Ethical aspects
Ethics can be described as critical and theoretical reflections of moral issues. In an academic context, we act with ethical considerations with respect to the university, other researchers and their work. This by referencing their writings and acknowledge their research. Bryman & Bell (2015, p. 134-145) describes the ethical research process with respect to individuals. Basic ethical principles regard integrity, voluntary participation in the research, anonymity and confidentiality. Ethical research also refers to privacy of the individuals (Bryman & Bell, 2015, p. 143). Ethical rules for conducting business regards the consent requirement which implies that participants decide over their own participation in the research, the information requirement which states that the participants should be well informed about the research purpose, and the useful
requirement which says that collected information is only intended to be used for the research purpose. Further, no participants should be harmed by the research. The researcher should neither provide the research participants false predictions or leak personal information (Bryman & Bell, 2015, p. 131). However, in this thesis we have not used surveys or interviews. There is no personal information in this research since this is a quantitative study built on secondary data and public corporate information. The authors own personal values will not affect the data. Further, the research is to the best of our abilities free from subjective interpretations, as we are doing our best to analyzing it based on an objectivistic view.

Since we have written this thesis at Umeå School of Business, Economics & Statistics and represent the university as students and ourselves, we find it highly important to conduct this research in an ethical and respectful way, based on high academic standards. Due to this, the study tries to follow the general recommendations stated by the Swedish Research Council concerning good research practice (Swedish Research Council, 2017, p. 10). These recommendations include to be truthful and open about the results and methodology. The authors of this study have tried to describe the methodological process thoroughly to ensure this. Further, the discussion and interpretation of the results have been based on earlier studies and in relation to respected financial theories. This study has also tried to inform the reader when discussions are based on assumptions, as well as when they are in a more hypothetical sense.

7.5 Societal implications
There is a growing interest of investing based on ESG criteria in the financial markets, both globally (GSIA, 2016, p. 7) and in Europe (Eurosif, 2018). This could be believed to derive from two different reasons, or a combination of the two. One, investors believe that it is a way to earn better returns. Or two, investors have other motives than simple financial ones when investing their funds. The findings in this study can only to some extent justify the first reason, but gives no indication to the opposite being true. However, the findings justify the second reasons with empirical evidence against a potential prejudice that the reason would have a financial trade-off. This goes back to the problem background and the long-lived debate of whether a firm can do well by also doing good, in terms of financial objectives and societal considerations respectively. The societal implications of our research are that firms can accomplish this, at least on the capital market. Finally, our belief is that research in this spectrum is highly beneficial and in the interest of the society, to promote a better world to live in, without compromising for future generations.
8. TRUTH CRITERIA

The final chapter will discuss the validity, reliability and generalizability of this study. This concerns the measurements taken to assure that these criteria are met, as well as critique to give the reader information about possible drawbacks of the study.

8.1 Reliability

Reliability in research concerns the precision and consistence of measurements. (Bryman & Bell, 2015, p. 169). One part of reliability is stability, which refers to whether a measurement is stable over other time. If measurements are stable, it should translate into the possibility of replicating a study and receiving the same results. This is of great importance, especially concerning quantitative research.

Concerning this study, most of the data used consist of publically available information about companies and financial markets, gathered as secondary data. That is, stock returns for individual companies, the stock returns that constitutes the factor models, the risk-free rate for the Eurozone as well as countries market capitalization. This is objective measures that leaves little room for interpretation. Further, they will not change over time, and would most likely be the same if one were to collect them from different sources. Even though the factor models are constructed by the provider, in this case AQR, one could replicate this based on publically available data and thereby receive the same factors. The other part of the data in this study concerns the ESG score from Thomson Reuters. The procedure and eventual concerns regarding this has already been described in the methodological chapter under section 5.4. Based on that, there is a risk that the ESG score would change slightly over time, based on the methodology of Thomson Reuters. This was something that the authors of this thesis learned during the data process. Due to this, we have as mentioned before stated that the data is available by the authors upon request. ESG scores could also differ based on providers (Chatterji et al., 2015). Thereby, replicating this study based on another provider of ESG might produce some differences in the results. In addition to this, whenever adjustments or own calculations based on data have been conducted, the authors of this study have tried their best to describe the procedure in order to enhance the reliability and the replicability.

Another aspect that can affect the reliability of our study is human error in the data processing and in the calculations. The data processes have been time consuming and demanding, and there is always the possibility of some errors occurring. However, preventive measure against this have been taken along the study. Many calculations have been double-checked and recalculated from the original data, based on random spot checks of the sample. This should satisfy the reliability in terms of minimizing the human error in the data process.

Another part of reliability is called internal reliability (Bryman & Bell, 2015, p. 169). This refers to if a measure is consistent within itself. In other words, for example how well the individual ESG metrics and measurement are related to each other, and then to the total ESG score. The data should be measured and calculated in the same way, with the same underlying factors to the ranking scale. However, this could be questioned since all data is first collected from company information, by over 150 analysts all around the globe, and then calculated. The range is huge, and one cannot be completely sure that the data is treated and exactly coherent with the metrics. However, since Thomson Reuters Eikon is a widely accepted source for financial data by investors, companies and the
academic society, and has standardized principles for business information and ESG calculations worldwide, the measures should be somewhat consistent.

8.2 Validity
Qualitative and quantitative research methods have different truth criteria to consider. However, Bryman & Bell (2015, p. 50) claims that validity is one of the most important research criteria among all different social scientific methods. Validity describes how relationships and generated results are coherent with each other. Construct validity, or measurement validity, refers to if measuring something really represent the concept that it is supposed to measure. That is, the underlying methods of the measurement in the study needs to be reliable (Bryman & Bell, 2015, p. 50). For this study, this concerns the measurement of ESG performance in form of the Thomson Reuters ESG score and the measurement of risk-adjusted returns in form of the Carhart four-factor method. The data process of Thomson Reuters covers over 400 ESG measures that is manually processed by over 150 research analysts, who are specifically trained in ESG related matters (Thomson Reuters, p. 4). This gives strength to the argument that the ESG score of Thomson Reuters is complex enough, and that the research analysts are equipped enough, to measure and rank companies CSR activities. A description of Thomson Reuters ESG score has already been covered in section 5.4, and we see no further reason to questioning the validity of the score. As for risk-adjusted returns, this study uses a performance measurement that are commonly used in earlier literature and is an established measurement of portfolio performance overall (Bodie et al. 2014, p. 433). As this study uses a sample only from the Eurozone, the authors of this thesis have further tried to assure the validity by constructing our own factors for this region specifically, as described in section 5.7.

Internal validity refers to the matter of causality in a study. That is, if the independent variable has a true and real effect on the dependent variable, or if the effect is explained by something that the researcher fails to account for (Bryman & Bell, 2015, p. 50). Since there are many variables that possibly could affect risk and return in financial markets, this is something that is very difficult to account for fully. In this study, valid relations through the results have been found and is presented based on the significance levels of 1%, 5% and 10%. However, this only proves a relation, not causality. The possible causal effect has been discussed in the analysis, the theoretical framework and in accordance with the hypothesis as well. The authors have also questioned the causality regarding some of the results. That is why we are careful to interpret the results of the significant best-effort governance portfolio based on a 20% cut-off rate, since the same governance portfolio based on a 10% cut-off rate showed insignificant results. Extern validity instead refer to the question of whether results are generalizable beyond the sample (Bryman & Bell, 2015, p. 50-51. The next section will cover this topic.

8.3 Generalizability
The term generalizability refers to which extent the results are applicable to other contexts and groups that are not incorporated in the researched sample (Bryman & Bell, 2015, p. 174). Representative selection of the data should be large enough to make sure that the results could be drawn to represent a population and not just the used sample. Bryman & Bell (2015, p. 176) claims that the results from the research only applies to the population, even though it might be tempting to generalize the results to a similar context.
This thesis focuses on the Eurozone, and more specifically the Euro STOXX 300 as an investable universe. Thus, the results of this index are the only thing we can argue strongly for. In a second stage, a strong argument of generalizability could be made for all companies on the financial markets in the Eurozone that has an ESG-coverage. As mentioned earlier, in the beginning of 2019, Thomson Reuters has ESG coverage of approximately 1200 companies in Europe (Thomson Reuters, 2019, p. 5). The complete coverage of the Eurozone is thereby smaller. Unfortunately, we have not been able to find the exact coverage for the Eurozone. Consequently, we have not been able to include all these in the sample. However, our investable universe of approximately 300 companies, where a clear majority have ESG coverage, should be enough to provide a generalizability over the Eurozone. Further, we believe that the results of this thesis could be generalizable to other markets, since the global financial markets today are highly interconnected. However, even if our results may not be limited to companies in our study, to draw the same conclusions for other markets and periods with certainty, one needs to conduct the same study under these specific settings.
REFERENCE LIST


https://data.worldbank.org/indicator/CM.MKT.LCAP.CD [Retrieved April 7, 2019]


Thomson Reuters. (2019). *Thomson Reuters Business Classification.* Available via:

UNFCC (2015). *Paris Agreement - Article 2.* Available via:


APPENDIX

Appendix 1. All hypotheses based on individual pillar scores

\begin{align*}
H_{04a} &: \alpha_{BE,Environmental} = 0 \\
H_{A4a} &: \alpha_{BE,Environmental} \neq 0 \\
H_{04b} &: \alpha_{BE,Social} = 0 \\
H_{A4b} &: \alpha_{BE,Social} \neq 0 \\
H_{04c} &: \alpha_{BE,Governance} = 0 \\
H_{A4c} &: \alpha_{BE,Governance} \neq 0 \\
H_{05a} &: \alpha_{BiC,Environmental} = 0 \\
H_{A5a} &: \alpha_{BiC,Environmental} \neq 0 \\
H_{05b} &: \alpha_{BiC,Social} = 0 \\
H_{A5b} &: \alpha_{BiC,Social} \neq 0 \\
H_{05c} &: \alpha_{BiC,Governance} = 0 \\
H_{A5c} &: \alpha_{BiC,Governance} \neq 0 \\
H_{06a} &: \alpha_{BE,Environmental} = \alpha_{BiC,Environmental} \\
H_{A6a} &: \alpha_{BE,Environmental} \neq \alpha_{BiC,Environmental} \\
H_{06b} &: \alpha_{BE,Social} = \alpha_{BiC,Social} \\
H_{A6b} &: \alpha_{BE,Social} \neq \alpha_{BiC,Social} \\
H_{06c} &: \alpha_{BE,Governance} = \alpha_{BiC,Governance} \\
H_{A6c} &: \alpha_{BE,Governance} \neq \alpha_{BiC,Governance}
\end{align*}

Appendix 2. Sectors and cut-off rates for portfolios

<table>
<thead>
<tr>
<th>ECONOMIC SECTOR</th>
<th>Euro Stoxx 300</th>
<th>20% cut-off</th>
<th>10% cut-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Materials</td>
<td>32,3</td>
<td>6,5</td>
<td>3,2</td>
</tr>
<tr>
<td>Consumer Cyclicals</td>
<td>44,7</td>
<td>8,9</td>
<td>4,5</td>
</tr>
<tr>
<td>Consumer Non-Cyclicals</td>
<td>22,7</td>
<td>4,5</td>
<td>2,3</td>
</tr>
<tr>
<td>Energy</td>
<td>19,0</td>
<td>3,8</td>
<td>1,9</td>
</tr>
<tr>
<td>Financials</td>
<td>63,7</td>
<td>12,7</td>
<td>6,4</td>
</tr>
<tr>
<td>Healthcare</td>
<td>16,7</td>
<td>3,3</td>
<td>1,7</td>
</tr>
<tr>
<td>Industrials</td>
<td>58,4</td>
<td>11,7</td>
<td>5,8</td>
</tr>
<tr>
<td>Technology</td>
<td>15,9</td>
<td>3,2</td>
<td>1,6</td>
</tr>
<tr>
<td>Telecommunications Services</td>
<td>16,8</td>
<td>3,4</td>
<td>1,7</td>
</tr>
<tr>
<td>Utilities</td>
<td>16,6</td>
<td>3,3</td>
<td>1,7</td>
</tr>
<tr>
<td><strong>TOTAL CONSTITUENTS</strong></td>
<td><strong>306,8</strong></td>
<td><strong>61,4</strong></td>
<td><strong>30,7</strong></td>
</tr>
</tbody>
</table>

This table presents the average constituents in each of the economic sector for Euro STOXX 300 during 2009-2018. The numbers are based on the constituents list at the 1st of January each year. Economic sector is based on Thomson Reuters Business Classification. Cut-off rates is the average amount of companies from each sector included in the portfolios over the whole period.
Appendix 3. Regressions for difference portfolios

```
. regress BEvsBICESG20 MKTRF SMB HMLFF UMD, vce(robust)

Linear regression                                   Number of obs  =    120
F(4, 115)     =    9.38
Prob > F      = 0.0000
R-squared     = 0.2328
Root MSE      = 0.01271

BEvsBICESG20 Coef. Std. Err. t P>|t| [95% Conf. Interval]
----------------- ------------- ----------- ------- -----------------  ------------------
MKTRF  -.0360766  .0340878 -1.03  0.303  -0.1051628  .0330096
SMB    .2129883   .0424176  5.02  0.000   .1289673   .2970093
HMLFF  -.2184844  .0517058 -4.23  0.000  -.3209037  -.1160651
UMD   -.0150011   .0267377 -0.56  0.576  -.0679633  .0379613
_cons  .0002761   .0011895  0.23  0.817  -.0020822  .0026322

. regress BEvsBICESG20 MKTRF SMB HMLFF UMD, vce(robust)

Linear regression                                   Number of obs  =    120
F(4, 115)     =    4.79
Prob > F      = 0.0013
R-squared     = 0.1577
Root MSE      = 0.01833

BEvsBICESG10 Coef. Std. Err. t P>|t| [95% Conf. Interval]
----------------- ------------- ----------- ------- -----------------  ------------------
MKTRF  -.0354426  .0515193 -0.69  0.493  -.1374924  .0666072
SMB    .277814    .0725183  3.83  0.000   .1341692   .4214587
HMLFF  -.1723591  .0799595 -2.16  0.033  -.3307424  -.0139758
UMD   -.0390311   .0405519 -0.96  0.338  -.1379566  .0698944
_cons  .001316    .0011895  0.78  0.439  -.0020401  .0046722
```
| BEvsBICEn~20 | Coef. | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|-------------|-------|-----------|------|------|---------------------|
| MKTRF       | .0031439 | .0352922 | 0.09 | 0.929 | -0.0667631          | .0730509 |
| SMB         | .1755301 | .0554481 | 3.17 | 0.002 | .065698             | .2853621 |
| HMLFF       | -.1861065 | .0526557 | -3.53 | 0.001 | -.2904073           | -.0818057 |
| UMD         | .0263585 | .0279398 | 0.94 | 0.347 | -.0289848           | .0817018 |
| _cons       | -.001006 | .001148  | -0.09 | 0.930 | -.0023745           | .0021733 |

| BEvsBICEn~10 | Coef. | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|-------------|-------|-----------|------|------|---------------------|
| MKTRF       | -.0721379 | .0431587 | -1.67 | 0.097 | -.157627            | .0033512 |
| SMB         | .2430857 | .0635078 | 3.84 | 0.000 | .1180888            | .3696825 |
| HMLFF       | -.188193  | .0472428 | -4.00 | 0.000 | -.3152173           | -.0662750 |
| UMD         | -.1025322 | .0338619 | -3.00 | 0.000 | -.2904073           | -.0818057 |
| _cons       | .0006774  | .0010979 | 0.62 | 0.538 | -.0014973           | .0028521 |

| BEvsBICSoc~20 | Coef. | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|--------------|-------|-----------|------|------|---------------------|
| MKTRF        | -.0046049 | .038763 | -0.12 | 0.906 | -.0813688           | .0721771 |
| SMB          | .2220003  | .0480661 | 4.62 | 0.000 | .1267905            | .31721   |
| HMLFF        | -.188193  | .0472428 | -3.98 | 0.000 | -.2817718           | -.0946142 |
| UMD          | -.0452572 | .0264785 | -1.71 | 0.090 | -.0977061           | .0071917 |
| _cons        | .0006774  | .0010979 | 0.62 | 0.538 | -.0014973           | .0028521 |
### . regress BEvsBICSocial10 MKTRF SMB HMLFF UMD, vce(robust)

Linear regression

| Coef.   | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|---------|-----------|-------|-------|---------------------|
| MKTRF   | -0.0306973 | 0.0511227 | -0.59 | 0.559   | -0.1319615, 0.0705669 |
| SMB     | 0.1875009   | 0.0631711 | 2.97  | 0.004   | 0.062371, 0.3126308   |
| HMLFF   | -0.1661084   | 0.0653032 | -2.54 | 0.012   | -0.2954615, -0.0367554 |
| UMD     | -0.0048579   | 0.0367986 | -1.36 | 0.177   | -0.1377489, 0.008033  |
| _cons   | 0.0013401    | 0.0014687 | 0.91  | 0.363   | -0.0015690, 0.0042492 |

Root MSE = 0.01617
R-squared = 0.1150
Prob > F = 0.0006
F(4, 115) = 5.27
Number of obs = 120

### . regress BEvsBICGovernance20 MKTRF SMB HMLFF UMD, vce(robust)

Linear regression

| Coef.   | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|---------|-----------|-------|-------|---------------------|
| MKTRF   | -0.0496329 | 0.0331922 | -1.50 | 0.138   | -0.1153803, 0.0161144 |
| SMB     | 0.0985325   | 0.0447579 | 2.20  | 0.030   | 0.0098758, 0.1871892  |
| HMLFF   | -0.1189035   | 0.0595349 | -2.00 | 0.048   | -0.2368308, -0.009763  |
| UMD     | 0.0455426    | 0.0419158 | 1.09  | 0.280   | -0.0374844, 0.1285697 |
| _cons   | 0.0002251    | 0.001343  | 0.17  | 0.867   | -0.0024351, 0.0028853 |

Root MSE = 0.01273
R-squared = 0.1824
Prob > F = 0.0011
F(4, 115) = 4.90
Number of obs = 120

### . regress BEvsBICGovernance20 MKTRF SMB HMLFF UMD, vce(robust)

Linear regression

| Coef.   | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|---------|-----------|-------|-------|---------------------|
| MKTRF   | -0.0496329 | 0.0331922 | -1.50 | 0.138   | -0.1153803, 0.0161144 |
| SMB     | 0.0985325   | 0.0447579 | 2.20  | 0.030   | 0.0098758, 0.1871892  |
| HMLFF   | -0.1189035   | 0.0595349 | -2.00 | 0.048   | -0.2368308, -0.009763  |
| UMD     | 0.0455426    | 0.0419158 | 1.09  | 0.280   | -0.0374844, 0.1285697 |
| _cons   | 0.0002251    | 0.001343  | 0.17  | 0.867   | -0.0024351, 0.0028853 |

Root MSE = 0.01273
R-squared = 0.1824
Prob > F = 0.0011
F(4, 115) = 4.90
Number of obs = 120