



**UNIVERSITY OF GOTHENBURG**  
**SCHOOL OF BUSINESS, ECONOMICS AND LAW**

# **EXPLORING THE RELATIONSHIP BETWEEN ESG DISAGREEMENT AND STOCK RETURNS**

**AN EMPIRICAL ANALYSIS**

Andrea Pasquali

Supervisor: Taylan Mavruk

Master's thesis in Finance

Spring 2023

Graduate School, School of Business, Economics and Law, University of Gothenburg,  
Sweden

Key words: ESG, Sustainability, portfolio analysis, rating disagreement

## Abstract

During the recent years investors shifted their preferences towards sustainable stocks and funds, increasing the importance of Environmental, Social and Governance ratings. At a similar pace the related literature started to develop shedding light on some crucial aspects of the ESG ratings, such as the disagreement between rating agencies and the lack of common methodologies to assess the ratings. This study makes a step forward, breaking down the relation between ESG ratings disagreement and stock returns. With only two datasets at disposal, provided by Morningstar and Refinitiv, the disagreement measure between the ratings is represented by the difference in absolute value between the ratings. After sorting the stocks each month based on the disagreement measure, a portfolio and stock analysis is conducted. The Fama and French 5-factor model and a two-sample t-test are employed to examine the existence of abnormal returns in the context of ESG disagreement. The results suggest that there is no advantage in investing in stocks with higher ESG rating disagreement; however, investing in a portfolio based on stocks with lower ESG disagreement will yield higher cumulative returns.

## Acknowledgments

I wish to express my gratitude to my supervisor Taylan Mavruk, who always welcomed me with a smile and provided me with essential insights and suggestions to improve my thesis. Furthermore, I express my gratitude to professor Rocco Ciciretti, from Italy, who followed the development of this study remotely.

# TABLE OF CONTENTS

<b>1. INTRODUCTION</b> .....	5
1.1. Introduction.....	5
1.2. Hypothesis.....	8
1.3. ESG rating definition .....	8
<b>2. LITERATURE REVIEW</b> .....	10
2.1. Previous Research .....	10
<b>3. DATA AND METHODOLOGY</b> .....	16
3.1. Sample description.....	16
3.1.1 Time Frame .....	16
3.1.2 Dropout Analysis.....	17
3.2 Variable Specification .....	18
3.2.1 Risk-free rate.....	18
3.2.2 Fama and French 5-factor model .....	20
3.3. Portfolio analysis.....	21
3.4. Two sample t-test.....	22
<b>4. EMPIRICAL RESULTS</b> .....	24
4.1. Baseline results.....	24
4.2. Discussions.....	30
4.3 Robustness analysis.....	31
4.3.1 ESG conversion.....	31
4.3.2 ESG rating sources .....	32
4.3.3. Portfolio selection .....	33
<b>5. CONCLUSIONS</b> .....	34
5.1. Further research.....	35
<b>References</b> .....	37
<b>Appendix</b> .....	40

# 1. INTRODUCTION

## 1.1. Introduction

Institutions that provide ratings for Environmental, Social, and Governance (ESG) factors have gained significant influence in the recent years. ESG assets surpassed \$35 trillion in 2020, up from \$30.6 trillion in 2018 and \$22.8 trillion in 2016, to become a third of the total global assets under management, according to the Global Sustainable Investment Association, while asset managers globally are expected to increase their ESG related assets under management to US\$33.9tn by 2026 (PwC, 2022). This increasing trend implies higher importance and responsibility for rating agencies, being the only reliable source for investors who seek sustainable and responsible investments. At the same time many academics are delving into the ESG domain, relying on the ratings to support their studies. Therefore, ESG ratings started to gain significant importance since they can influence investors' decisions and shape the direction of capital allocation.

Despite the important role these ratings play, there is a substantial divergence between ratings provided by different agencies. Different studies measured the average correlation between ESG ratings, yielding to similar results, such as a correlation of 0.48 (Avramov, Cheng, Lioui, & Tarelli, 2022), or between 0.38 and 0.71 (Berg, F. Kolbel, & Rigobon, 2022).

The proficient literature on this topic arose the question whether there is a relation between the rating disagreement and stock returns. Even though there are more than 1000 papers written between 2015 and 2020 (Whelan, Atz, Van Holt, & Clark, 2021) regarding the relation between stock returns and ESG ratings, little attention has been devoted on how uncertainty in ESG ratings affects asset prices, leaving room for further researches.

The aim of this paper is to investigate whether there is a performance trade off in choosing portfolios and stocks that exhibit high ESG disagreement, using recent data provided by two important raters, Morningstar and Refinitiv.

Brandon, Krueger and Schmidt (2021) bring some evidence to this subject studying whether stock returns are related to ESG disagreement in the US market, finding out a positive relation. However, their analysis is strictly related to the US market, limiting the extent to which this study can be applied. I extend previous research by shifting the focus to the European market, considering the Euro Stoxx 600 index that regroups the 600 Blue Chip

stocks of countries included in the Eurozone. By shifting the focus from the United States to Europe, this research not only seeks to address a gap in the existing literature but also aims to present findings pertaining to a market that has been experiencing significant growth from the perspective of environmental, social, and governance (ESG) considerations. Europe has emerged as a key player in the realm of ESG, with increasing investments being directed towards sustainable and socially responsible practices. By 2025, it is projected that the value of ESG assets domiciled in Europe will range from €7.4 trillion to €9 trillion. This indicates a substantial growth in ESG investments within the region. Additionally, ESG funds are expected to represent a significant portion of total European Mutual Fund assets, between 46% and 56% by 2025, compared to only 37% at the end of 2021. Furthermore, there is a strong trend among European institutional investors towards ESG investments. A significant 66% of these investors have plans to cease investing in non-ESG funds, with an additional two-thirds expected to do so by the end of 2023 (Funds-Europe, 2023). This suggests a significant shift in investor preferences towards more sustainable and socially responsible investment options in Europe. This paper is the first one to test whether the disagreement about a firm's ESG performance has consequences on stock returns in the European market.

The lack of interest for this topic lasted for many years, since investment decisions were driven by only one purpose: gaining the highest possible return. Nobel Prize in Economic Sciences Milton Friedman wrote in the New York Times Magazine, in 1970, that “*there is one and only one social responsibility of business — to use its resources and engage in activities designed to increase its profits*”, pointing out that the social responsibility that goes beyond serving the interests of stockholders is just a misconception of the nature of a free economy (Friedman, 1970). Even though this monetarist view has been the driving force of financial investments for many years, the trend is now changing. Investors started to care about the impacts of their investments outside the financial world, taking into considerations Environmental, Social and Corporate Governance (ESG) aspects in their investment decisions. Part of this change is driven by investment firms that are increasing their volume, such that modern portfolio theory can not be used to mitigate system-level risk: “*firms with trillions of dollars under management have no hedge against the global economy, i.e., they have become too big to let the planet fail*” (Eccles & Klimenko, 2019). In addition, the rise of social media has had a profound impact on both individuals and businesses. People now have the ability to use their smartphones to shine a light on unethical practices, putting pressure on companies to be more vigilant in their actions to avoid negative publicity and

reputational harm. At the same time, social media allow the spread of information in real time all around the globe increasing the involvement, both physical and emotional, of every person in these events. It all increased the awareness of the effects of human's actions on the environment, on the biodiversity in and outside of the working place and on the gender and race equality. A recent example of how social media shed light on unethical practices is represented by Shell, that promised to invest 6 bn USD in green between 2016 and 2020 without keeping up the promise (Rinnovabili.it, 2020). Global Witness accused Shell, highlighting that only 1.5% of Shell's spending was used for renewable solutions, compared to the 12% claimed by the company. Zorka Milin, senior advisor at Global Witness, defined Shell's renewable policies as "*pure fictions*" (Global Witness, 2023). Sources like The Economist, Washingtonpost and Euronews denounced Shell on different platforms, facilitating the spread of the information.

For this reason, investors started to incorporate these aspects in their investment strategies, looking for firms that take into account social, ethical and environmental decisions.

Investments in ESG assets increased and, consequently, the demand for ratings that could summarise in a single metric the overall ESG position of a firm. However, the response to this increasing demand is insufficiently regulated and not homogeneous: rating agencies are independent third parties that implemented their own rating techniques and criteria, resulting often in a low correlation between ratings. Brandon et al. reported that the average pairwise correlation between ESG ratings of seven rating providers, for firms in the S&P500, is about 0.45 (Brandon, Krueger, & Schmidt, 2021). This lack of regulation and homogeneity in defining how ESG ratings should be calculated works as barrier in this developing environment, where investors can not completely rely on rating agencies, increasing the probability of being involved in greenwashing scandals; at the same time, firms that want to increase their rating will struggle given that every rating agency uses different metrics.

The paper is organized in the following way: after defining the hypothesis below, chapter two provides an overview of the previous research on this topic and points out results that will be useful for the discussion. Chapter three focuses on the data and methodology of the study, providing information about the sample, the variables and the portfolios. Chapter four reports the main findings of this paper and the related implications.

To conclude, chapter five points out the conclusions and possible future research on this topic.

## 1.2. Hypothesis

The ESG sphere poses a huge challenge for rated firms. The empirical results of this paper help to reduce the uncertainty and better understand the effects of the dispersion of ratings on stock returns. To do that, I select the stocks with highest a lowest rating disagreement and investigate through an econometric model if there exists a relation between their returns and the disagreement in ESG ratings. More specifically, these two hypotheses are tested:

Hypothesis one:

*H<sub>0</sub> : There is a performance trade-off in investing in a portfolio with high ESG disagreement*

*H<sub>1</sub> : There is not a performance trade-off in investing in a portfolio with high ESG disagreement*

Hypothesis two:

*H<sub>0</sub> : Stocks with high ESG disagreement exhibit higher abnormal returns*

*H<sub>1</sub> : Stocks with high ESG disagreement do not exhibit higher abnormal returns*

## 1.3. ESG rating definition

The ESG rating is used by investors to assess the sustainability and ethical practices of companies or investment funds. ESG stands for Environmental, Social, and Governance, and the rating system evaluates a company's performance on these three factors. The environmental factor considers a company's impact on the environment and climate change, taking into consideration aspects like CO<sub>2</sub> emissions reduction and waste generated. The social factor looks at how the company treats its employees and its impact on the communities it operates in, as well as gender equality and biodiversity in the working place. The governance factor assesses the company's leadership and management practices, including issues like transparency and ethical behaviour, in order to reduce corruption and misbehaving.

ESG ratings are issued by different institutions such as banks, rating agencies, research firms and non-profit organisations. Each entity has its own distinct rating system, often represented in different ways, such as numerical scores or letter grades, and on different scales. Two



examples of rating agencies that provide ESG ratings are Refinitiv and Morningstar. These independent companies offer a range of measures, including ESG ratings, to help investors assess the sustainability of a company. In this paper, ESG ratings from Refinitiv and Morningstar are used as a measure of sustainability.

Refinitiv captures and calculates over 630 company-level ESG measures, grouped in 10 categories and then rolled up into the three pillar scores E,S,G. Their ratings range from 0 to 100 and are supported by letter grades (from A+ to D-) to show at a glance how companies are performing (Refinitiv, 2022). The higher is the grade/score, the higher is the ESG performance of the company.

Morningstar, rating over 1500 companies worldwide (Morningstar, 2021), kept a 0-100 scale for the ratings until September 2019. After, the scale ranges from 0-50 where lower ratings represent lower ESG-related risk.

## 2. LITERATURE REVIEW

### 2.1. Previous Research

Environmental, Social, and Governance (ESG) has emerged as a prominent framework in the realm of responsible investing and corporate sustainability. The term ESG was officially coined in 2004 with the publication of the UN Global Compact Initiative's report "Who Cares Wins" (Who Cares Wins, 2004), with the ambitious purpose to summarize in one value the main ethical financial pillars. After that, the United Nations Principles for Responsible Investment (PRI) were introduced in 2006, comprehending 734 signatories in 2010, 1.384 in 2015 and 3.038 in 2020 (PRI, 2021). The PRI is based on six Principles aiming to promote the incorporation of ESG factors in investment strategies and each signatory represents a commitment to this goal. To truly understand the essence and meaning of ESG, it is crucial to examine its roots in the concepts of Corporate Social Responsibility (CSR) and Socially Responsible Investing (SRI), which have played a pivotal role in shaping this framework.

Formal writings on CSR are mostly a product of the second half of the 20<sup>th</sup> century, mainly in the United States. This concept was initially referred to as Social Responsibility (SR), as described many years before by Howard R. Bowen in his book "Social Responsibility of the Businessman" (1953), widely considered the beginning of the literature on this topic. Bowen, sometimes considered the *father of Corporate Social Responsibility* (Carroll A. , 1999), points out that the largest firms have a huge influence on many aspects of life that are not directly related with their business, highlighting the relation between the firm's operations and the impacts they have on the surrounding environment. In his book Bowen defines Social Responsibility as "*the obligations of businessmen to pursue those policies, to make those decisions, or to follow those lines of action which are desirable in terms of the objectives and values of our society*" (Bowen, 1953).

The concept of SR kept evolving during the 60s, a decade in which more definitions were provided by several authors. Keith Devis defines SR as the "*businessmen's decisions and actions taken for reasons at least partially beyond the firm's direct economic or technical interest*" (Davis, 1960); Clarence C. Walton defines it as the concept that "*recognizes the intimacy of the relationships between the corporation and society and realizes that such relationships must be kept in mind by top managers as the corporation and the related groups pursue their respective goals*" (Walton, 1967). Definitions and papers about SR, that started

to be referred to as Corporate Social Responsibility, proliferated during the 70s, allowing the studies during the 80s to be more focused on research and empirical analysis about this topic: for example, Cochran and Robert Wood (1984) and later on Aupperle, Carroll, and Hatfield (1985) studied the relationship between CSR and profitability.

At the beginning of the 90s one of the most common and established theories about CSR's structure was proposed by Archie B. Carroll. He modelled the CSR structure as a pyramid, with the economy category as the base, built upward through legal, ethical and philanthropic categories, pointing out that the business should not fulfil them in a sequential order, but all at the same time; more precisely, "*the CSR firm should strive to make a profit, obey the law, be ethical, and be a good corporate citizen*" (Carroll A. B., 1991).

In the last two decades the CSR kept evolving, shifting from minimizing local harm to tackling global issues. Today, CSR refers to the voluntary actions and initiatives undertaken by companies to integrate social and environmental concerns into their business operations. It goes beyond legal and regulatory requirements and emphasizes the role of businesses in addressing societal challenges and promoting sustainable development. Even if it is not mandatory to disclose information about responsible actions, firms' managers recognised the impact that this may have on costumers. For this reason, the motivations behind CSR vary among companies. Some embrace CSR as a moral imperative, driven by the desire to make a positive impact and contribute to society. Others recognize the business benefits of CSR, such as enhanced reputation, increased customer loyalty and attracting and retaining talented employees. Nearly 90 percent of companies in the S&P500 disclosed a CSR report in 2019, compared to only 20 percent in 2011 (Stobierski, 2021).

Therefore, CSR refers to the firm's voluntary initiatives and activities aimed at positively impacting society and the environment. The voluntary disclosure about CSR represents the supply of sustainable information for investors. On the other hand, the demand is represented by Socially Responsible Investing (SRI). SRI, also known as sustainable investing or ethical investing, represents an investment approach that considers both financial returns and environmental, social and governance aspects, aiming to promote the development of ethical practices and fund those companies that care about their impact on the surrounding environment. As CSR, SRI is not a recent concept. Examples of SRI can be dated back to 1758, when the Religious Society of Friends prohibited members from participating in the slave trade. Focusing on more recent years, the modern era of SRI investments evolved during the 60s, as a reaction to some relevant geopolitical events and some unethical

practices that were systematically carried on all over the world. During these years the Vietnam War was ongoing, raising general dissent among investors of companies involved in war-related sectors: we may recall that in that period the wide use of Napalm spurred several people to protest against Dow Chemical Company and disinvest from companies that were benefitting from the war. In the same period, Apartheid was enacted in South Africa: also in this case, the social awareness of those events pushed investors and big companies to reallocate their investments, disinvesting from companies operating in South Africa. The common knowledge about this topic and the pressures arisen by the citizens even eventually led the United Nations to impose an arms embargo against South Africa. Following these events, in 1971 the first socially responsible mutual fund was launched in the US by Pax World.

This long phase of reaction to a war considered unequal by a part of the population was one of the core components of SRI, that all considered, consists in the allocation of funds in projects and investments that are perceived by the investors as worthy to promote equality, sustainability or ethic principles.

In the last decades the concern for SRI experienced a considerable surge of interest as a consequence of extreme weather events and of the increasing awareness of social problems and inequalities that are still present around the world (BLM<sup>1</sup>, gender equality, etc..). In the context where the supply is represented by the CSR and the demand by the SRI, the Regulators stepped in trying to define a measure that could help investors to discriminate between what can and what cannot be considered a socially responsible investment. For this reason, standard measures were introduced to assess the firm's incorporation of Environmental, Social, and Governance (ESG) aspects in its operations. After the introduction of ESG ratings, ESG asset under management (AuM) increased in volume year by year and are expected to grow even more in the future: asset managers globally are expected to increase their ESG-related assets under management to US \$33.9tn by 2026, from US \$18.4tn in 2021 (PwC, 2022).

The increasing number of investors who rely on independent third parties to assess firms' ESG performances has resulted in rating agencies gaining significant influence over the last two decades. These agencies enable investors to differentiate between firms on the basis of their ESG performance, much like credit rating agencies differentiate based on

---

<sup>1</sup> Black Lives Matter

creditworthiness. However, according to Berg et al. (2022), there are at least three important differences between the two. Firstly, while creditworthiness is well defined and represented by the probability of default, the concept of ESG is not well defined yet and part of the ESG score still relies on how the raters interpretate the concept of ESG performance and how they decide to measure it. Secondly, ESG reporting is still in its early stages, resulting in various reporting standards for ESG disclosures, many of which are voluntary, and different assessment methods. Lastly, ESG raters are paid by investors rather than the rated companies, which eliminates the problem of the so called “rating shopping” seen with credit raters (Berg, F. Kolbel, & Rigobon, 2022).

The most important differences are the first two, which explain why credit ratings exhibit a 99% correlation, while ESG ratings do not. Various studies report different levels of correlation between ESG ratings, with Avramov et al. (2022), reporting an average correlation of 0.48, Brandon et al. (2021) reporting an average correlation of 0.45, and Berg et al. (2022) reporting correlations ranging from 0.38 to 0.71. All of these studies analyse ratings from a range of prominent ESG raters to obtain their results. The lack of coherence among these ratings poses several problems, including discouraging companies from improving their ESG scores, as they may only be recognized by certain raters. Additionally, it is challenging for investors to identify sustainable products, increasing the risk of greenwashing, which can lead to reputational risks and discourage investments in ESG products. To address this issue, various studies have attempted to find solutions.

In their 2020 study, Billio et al. investigate the convergence of raters' methodologies and the consistency of the ratings they provide. The researchers believe that the primary difference in ratings is related to the data source and how the data are processed, as rating agencies use different sources of information and assessed indicators. For example, MSCI and FTSE Russel assess respectively 37 and 300 ESG criteria. Basing their analysis on four representative ESG indexes, they conclude that the overlap of common constituents is approximately 15%. They also find that the lack of common standards for ESG makes it difficult to evaluate the sustainability of a company, leading to some cases being deemed "*unratable*" (Billio, Costola, Hristova, Latino, & Pelizzon, 2020).

In 2022, Berg et al. conduct a study analysing data collected from six ESG rating agencies. They examine the ratings' divergence and identify the contributions of scope, measurement, and weight. Scope pertains to instances where ratings are based on different attribute sets,

measurement refers to situations where agencies measure the same attribute using different indicators, and weight considers the different attribute weights given by raters.

Berg et al. show that measurement accounted for 56% of the divergence, followed by scope at 38% and weight at 6%. These results indicate that the divergence is not only due to differing definitions, but it also reflects a fundamental disagreement concerning the underlying data (Berg, F. Kolbel, & Rigobon, 2022).

While part of the literature focuses on why there exists ESG disagreement, at the same time another branch of research investigates whether ESG ratings are related to stock returns. Friede, Busch and Bassen (2015) consider more than 2000 studies from 1970 to 2015 that focus on the relation between ESG criteria and financial performance, in order to unify the previous and prolific literature (Friede, Busch, & Bassen, 2015). The results show that roughly 90% of the studies find a non-negative relation between ESG and financial performance, and the large majority report a positive relation; in addition, they highlights that the positive impact of ESG on financial performance appears stable over time (Gunnar, Busch, & Bassen, 2015). Moreover, Whelan et al. (2020) examine the relationship between ESG and financial performance in more than 1000 papers from 2015 to 2020, finding a positive relation between ESG and financial performance for 58% of the corporate studies focused on operational metrics (ROA, ROE, stock performance), with 13% showing neutral impact, 21% mixed results and only 8% showing a negative relation (Whelan, Atz, Van Holt, & Clark, 2021).

The past literature suggests that two primary patterns have emerged: one investigates why there is disagreement in ratings, while the other explores whether there is a correlation between ESG scores and stock performance. Surprisingly, very few studies have examined the connection between ESG disagreement and stock performance.

Billio et al. (2020) prove that ESG disagreement disperses the effect of preferences of ESG investors on asset prices, to the point that even when there is agreement, the latter is so weak that it has no impact on the financial performances of ESG portfolios, stressing that financial performances would be different if all major rating agencies agreed on their rating methodologies, leading to a more homogeneous stock selection (Billio, Costola, Hristova, Latino, & Pelizzon, 2020). Their study is based on a wide group of companies included in the MSCI World Index, that takes into account 23 developed countries; the data used were collected by four different rating providers (Billio, Costola, Hristova, Latino, & Pelizzon, 2020).

Moreover, Brandon et al. (2021) delve into this topic focusing on the US market. They collect and study ESG ratings from seven prominent ESG rating providers for firms in the S&P 500 Index between 2010 and 2017. The disagreement measure is calculated as the standard deviation of the ratings provided for a firm at a given point in time. This measure is calculated on the overall ESG rating and, separately, on the E,S,G constituents. According to their results, stock returns are positively related to ESG rating disagreement; more precisely, this relation is mainly driven by the disagreement about the environmental rating. This positive relation is justified by the authors using a “*standard asset pricing argument*”: the ESG disagreement is perceived by investors as a source of risk, therefore requiring an uncertainty premium (Brandon, Krueger, & Schmidt, 2021).

This recent study, that offers some valuable results, is one of the few available on this topic. Considering that none of the above mentioned studies is uniquely focused on the European market, I exploit this gap in the literature and use it as an opportunity for new insights to be explored.

### **3. DATA AND METHODOLOGY**

#### **3.1. Sample description**

The reference sample for this study is represented by the STOXX Europe 600, which entails a fixed number of 600 components representing large, mid and small capitalization companies among the following 17 European countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and United Kingdom. The index covers approximately 90% of the free-float market capitalization of the European stock market (not limited to the Eurozone). For this reason, it is chosen as representative for the European market.

Furthermore, in order to mitigate the effects of survivorship bias, the sample includes all stocks that were part of the index during the reference period, regardless of whether they are still present or not.

All the data has been collected through Morningstar and Thomas Reuter DataStream (Refinitiv). While they are providing the same type of data, the availability and integrity of the two datasets is totally different. More precisely, Morningstar presents less complete data, often missing for a long period of time. Individual E, S and G ratings are not considered since they are available only on Refinitiv. In addition, a comparability issue has arisen in the analysis due to Morningstar's change in its rating system effective from September 1, 2019.

##### **3.1.1 Time Frame**

The report examines two different time periods. From 1<sup>st</sup> of January 2013 to 1<sup>st</sup> of September 2019 and from 1<sup>st</sup> of January 2013 to 1<sup>st</sup> of January 2023. This choice was made for several reasons. Firstly, prior to 2013, Morningstar's data were incomplete and contained several missing values, potentially introducing bias into the analysis. Secondly, Morningstar changed its rating system after September 2019, forcing the conversion of the new ratings in the previous scale, from 0 to 100, in order to allow the comparison with Refinitiv's data. For this reason, in order to avoid bias due to this manipulation of data, two different analysis are carried out.



### 3.1.2 Dropout Analysis

The presence of missing values in datasets is a common issue in empirical research, particularly when the data relates to companies that were either founded or have left the market during the interest period, i.e., 2013-2023. In the case of the STOXX 600 index, this issue is particularly relevant given that the index comprises a broad range of European companies from different sectors, some of which are likely to be newer or to have ceased operations in recent years. To deal with this issue, the study's methodology involves selecting the subset of stocks that have available values at the beginning of each month. By doing so, the analysis ensures that only stocks with complete information are used in the sample. Consequently, a subset of the original sample, namely the STOXX 600, is used. However, this technique still results in the exclusion of only 15 to 20 stocks for each month.

Moreover, some stocks were excluded from the dataset because of Morningstar's lack of integrity: some stocks' time series presented data for a period of time, then missing values for several years, and then data again. These stocks were not considered in this study.

To limit the effect of possibly spurious outliers a Winsorization (or Winsoring) is applied to the returns. It is the transformation of statistics by limiting extreme values in the statistical data. The procedure consists in setting all outliers equal to a specified percentile of the data. For example, a 90% winsorization would see all data below the 5<sup>th</sup> percentile set to the 5<sup>th</sup> percentile, and data above the 95<sup>th</sup> percentile set to the 95<sup>th</sup> percentile. Winsorized estimators are usually more robust to outliers than their more standard forms. In this study this procedure considered the cutoffs of 1% and 99%: all data below 1<sup>th</sup> percentile are set equal to the 1<sup>th</sup> percentile; all data above 99<sup>th</sup> percentile are set equal to the 99<sup>th</sup> percentile.

### 3.1.3 Morningstar's data after September 2019

After September 2019, Morningstar's rating scale changed from 0-100 to 0-50, where lower ratings represents lower ESG-related risk. The new rating scale can be interpreted in the following way:

- Overall Score of 0-9.99 points: Enterprise value is considered to have a negligible risk of material financial impacts driven by ESG factors

- Overall Score of 10-19.99 points: Enterprise value is considered to have a low risk of material financial impacts driven by ESG factors
- Overall Score of 20-29.99 points: Enterprise value is considered to have a medium risk of material financial impacts driven by ESG factors.
- Overall Score of 30-39.99 points: Enterprise value is considered to have a high risk of material financial impacts driven by ESG factors.
- Overall Score of 40 and higher points: Enterprise value is considered to have a severe risk of material financial impacts driven by ESG factors

The whole study depends on the availability of the disagreement measure between rating agencies that will represent the criteria for the stocks selection. The rating change in Morningstar's data poses a comparability problem, since the two datasets are expressed in different scales now. To overcome this issue, a manual conversion is performed on Morningstar's dataset after September 2019 to convert it back to a 0-100 scale. To do that, the following formula is applied:

$$ESG_{0-100} = 100 - 2 * ESG_{0-50} \quad \text{Equation 1}$$

$ESG_{0-100}$  represents the ESG rating converted in a 0-100 scale and  $ESG_{0-50}$  represents the ESG rating after September 2019.

To give an example, if  $ESG_{0-50} = 5$ , i.e. the risk is negligible, the corresponding  $ESG_{0-100}$  will be 90, corresponding to a high ESG score. More examples of how the conversion works are presented in the Appendix.

## 3.2 Variable Specification

### 3.2.1 Risk-free rate

The risk-free rate refers to the rate of return of an investment that carries a very low risk. In practice, it is often represented by the yield on a government bond of a country with a stable economy, such as the US Treasury bond. The risk-free rates used in this study are provided by Kenneth R. French Data Library (French, 2023) and represent the risk free rates related to Europe. It is calculated based on 16 countries, of which 15 included in the STOXX Europe

600. The time series of the risk-free rates will be used in Fama and French 5-factor model.

### 3.2.2 Disagreement Measure

Having just two datasets at disposal to compare the ESG ratings, the disagreement measure could not be represented by the correlation or the standard deviation, as previous studies do. For this reason, the disagreement measure is given by the absolute difference between the ratings:

$$Disagreement_{i,t} = | ESG_{i,t}^{ref} - ESG_{i,t}^{morn} | \quad \text{Equation 2}$$

$ESG_{i,t}^{ref}$  represents the ESG score of stock  $i$  at month  $t$  given by Refinitiv and  $ESG_{i,t}^{morn}$  the ESG score of stock  $i$  at month  $t$ , given by Morningstar

### 3.2.3 Descriptive Statistics

Table below reports the descriptive statistics of the ESG ratings provided by Refinitiv and Morningstar. The number of stocks is directly affected by the dropout analysis presented in section 4.1.3. The difference between the median and the mean in both samples implies that both distributions are left-skewed, i.e., most of the ratings are clustered at higher levels. This asymmetry in the distribution is shown by the two histograms reported below, representing the frequency distribution of the given ratings.

	Refinitiv	Morningstar
N. of stocks	585	585
Average ESG score	63.44	63.15
Percentile 25th	52	55
Median ESG score	66.54	64
Percentile 75th	77	73
Max ESG score	95.98	94
Min ESG score	2.6	3.66

*Table 1, Descriptive Statistics*

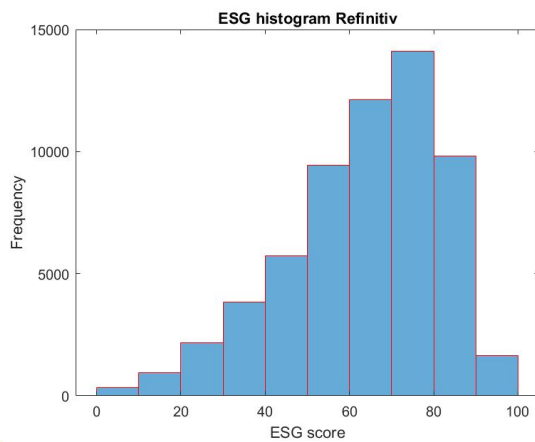


Figure 1, ESG Score Histogram for Refinitiv

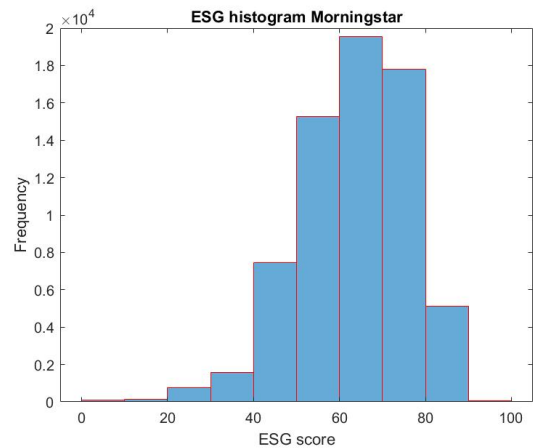


Figure 2, ESG Score Histogram for Morningstar

### 3.3. Method

For this study, a quantitative approach was selected, based on well-established econometric models and techniques. Sorting the stocks based on the disagreement in ESG ratings allowed the creation of portfolios<sup>2</sup> and enabled a performance comparison through the application of the Fama and French 5-factor model, consistent with previous literature. To test the second hypothesis of this study, i.e., the presence of abnormal returns in stocks with high ESG disagreement, Fama and French model is applied at a stock level and two-sample t-test is performed.

#### 3.2.2 Fama and French 5-factor model

Fama and French 5-Factor Model is an asset pricing model developed by Eugene Fama and Kenneth French in 2014 that aims to explain the relation between expected stock returns and risk. Several asset pricing models were proposed before this one, such as the Capital Asset Pricing Model (CAPM) in 1964, the Fama and French 3-Factor Model in 1992, and the Carhart Model in 1997. However, previous research have found that these models have limitations. As a result, the 5-factor model is an extension of the earlier Fama and French 3-Factor Model, obtained by adding two more explanatory variables to better capture the estimation of the expected returns. This model is widely used in the literature as reference model, given its explanatory power, and for this reason will be used in this study. According to Fama and French (2014), the model explains between 71% and 94% of the cross-section

<sup>2</sup> The analysis was conducted through the software Matlab

variance of expected returns for the Size, B/M, OP, and Inv. Portfolios (Fama & French, 2014).

The model is formulated as follows:

$$R_{it} - R_f = \alpha_i + \beta_{1,i} * (R_{mkt} - R_f) + \beta_{2,i} * SMB_t + \beta_{3,i} * HML_t + \beta_{4,i} * CMA_t + \beta_{5,i} * RMW_t + e_{it} \quad \text{Equation 3}$$

$R_{it}$  represents the expected portfolio return,  $R_f$  the risk free return,  $\beta_{1,i}$  the market risk,  $R_{mkt}$  the market portfolio return,  $\beta_{2,i}$  the size factor sensitivity,  $SMB_t$  the size factor,  $\beta_{3,i}$  the value factor sensitivity,  $HML_t$  the value factor,  $\beta_{4,i}$  the investment factor sensitivity,  $CMA_t$  the investment factor,  $\beta_{5,i}$  the profitability factor sensitivity,  $RMW_t$  the profitability factor and  $e_{it}$  the residuals of the model

The size factor SMB (small minus big) measures the excess returns of small-cap stocks over large-cap stocks; the value factor HML (high minus low) measures the excess returns of value stocks over growth stocks. The investment factor CMA (conservative minus aggressive) measures the excess returns of companies with conservative investment policies over those with aggressive ones. Finally, the profitability factor RMW (Robust minus weak) measures the excess returns of high profitability companies over low profitability ones.

All the factors are manually calculated using the available data, instead of retrieving them from Kenneth R. French Data Library, in order to present an in-sample analysis.

### 3.3. Portfolio analysis

Portfolio construction is commonly used to investigate the performance of ESG investing. This allows the application of basic asset-pricing models. In addition, it provides a straightforward and practical strategy for investors to exploit a potential relationship between ESG ratings and asset returns.

The portfolio construction in this paper is based on the following procedure that repeats in the same way each month: at month  $t$ , all the stocks are sorted according to their ESG disagreement measure, from smallest disagreement to largest. After, the two portfolios are built selecting the firms in the 30<sup>th</sup> and 70<sup>th</sup> – 100<sup>th</sup> percentile of the sorted distribution; the first portfolio will be called “agreement portfolio”, while the second one “disagreement

portfolio”. The monthly return of the portfolio is given by the weighted average of the returns of each single stock, where the weights are represented by the market capitalisation. Iterating this procedure for each month I obtain the monthly returns of the portfolio and use them as dependent variable in the Fama and French 5 factors model. This procedure is repeated for the two considered time frames.

### 3.4. Two sample t-test

After the portfolio analysis, a stock level analysis is considered: at the stock level, a two-sample t-test is used to examine whether stocks with high ESG disagreement exhibit significantly different abnormal returns when compared to those with low ESG disagreement. Under the null hypothesis, there is no difference between the means of the abnormal returns of the two samples. The following test statistic is applied:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{Equation 4}$$

t = test statistic

$\bar{x}_{1,2}$  = mean of sample 1,2

$n_{1,2}$  = number of observation of sample 1,2

$\sigma_{1,2}^2$  = variance of sample 1,2

This test statistic is used for the following system of hypothesis:

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

To conduct this analysis, the stocks are selected in the same way as for the portfolio analysis, but in this case the stock selection starts in January 2018. In this way for each stock there will be 60 observations available (from January 2013 to January 2018), necessary to run a time-series regression. After selecting the stocks in the 30<sup>th</sup> and 70<sup>th</sup> percentile as described before, Fama and French 5 factor model is applied to each stock in order to retrieve the values of alpha. This procedure is iterated for each month until the end of the sample. In the end, there will be two distributions of alpha, the first composed by the alphas related to the stocks present in the 30<sup>th</sup> percentile and the second of those in the 70<sup>th</sup> to 100<sup>th</sup> percentile.

The two sample t-test will test whether the means of these two distributions can be considered statistically equal or not.

Again, this procedure is repeated for the two considered time frames.

## 4. EMPIRICAL RESULTS

As previously mentioned, two different analysis are presented. One regarding the time period from January 2013 to September 2019, the date when Morningstar changed its rating system, and one considering the whole dataset, from January 2013 to January 2023. Both analysis are based on the same statistical approaches and methodologies.

### 4.1. Baseline results

#### January 2013 – September 2019 sample

First of all, figure 2 shows the cumulative returns of the agreement and disagreement portfolio constructed according to section 3.3. This first analysis provides evidence on the possible difference in choosing a strategy based on the ESG disagreement.

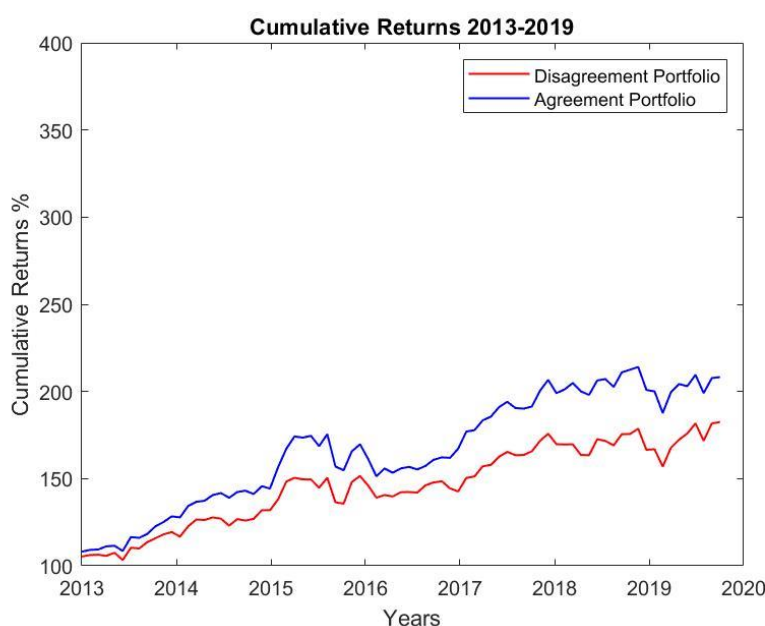


Figure 2, Cumulative Returns for the Agreement and Disagreement portfolio, 2013-2019

The cumulative returns are based on a time period of ca. 7 years, between 2013 and 2019. Both lines show a similar pattern, having peaks during the same periods. In fact, the correlation between the two cumulative returns is 0.992, implying that the agreement and disagreement portfolio tend to vary in tandem. However, the graph clearly shows that the



agreement portfolio had consistently a higher volume of cumulative returns throughout the whole period and their difference seems to increase with time. This simple analysis suggests that the agreement portfolio is better preferable than the disagreement one from the performance point of view.

Table 2 below reports the results of Fama and French 5 factor model applied to the portfolios. Having one return per month, the total number of observations is 79, as the months between January 2013 and September 2019. Both models report a high adjusted R-squared, implying that there is a good fit to the data. Regarding the coefficients of the regressions, the agreement portfolio presents a positive value of alpha while, on the contrary, the disagreement one presents a negative alpha. In the first case the value is not significant at a 5% level, given a p-value of 0.28, so it can not be considered statistically different from zero. In the second one the p-value is significant, equal to 0.014; however, its value is negative, so according to the model this strategy does not achieve any excess return. The value of beta is significant in both cases at 1% level. For the agreement portfolio it is slightly larger than one, while for the second is basically one. Since beta measures the sensitivity of the asset's returns to changes in the market returns, a beta value of 1 indicates that the asset's returns move in line with the market returns, while a beta value greater than 1 suggests that the asset is more volatile than the market. In our case, the disagreement portfolio is subject to a smaller systematic risk, given the smaller value of beta. The result is consistent with the findings in Monasterolo and De Angelis (2020) where the sustainable asset class shows a market exposition that is larger than the unsustainable one (Monasterolo & De Angelis, 2020), assuming a positive relation between the disagreement in the ratings and the unsustainability of the firm.

	Agreement Portfolio	Disagreement Portfolio
	Coefficients	Coefficients
Alpha	0.00114 (0.00105)	-0.00234** (0.00093)
Rm - Rf	1.01270*** (0.03114)	0.97822*** (0.02769)
SMB	-0.0934 (0.08135)	-0.10932 (0.07232)
HML	0.07043 (0.07892)	-0.09502 (0.07016)
RMW	0.04452 (0.05919)	-0.11072** (0.05263)
CMA	-0.00195 (0.06051)	0.01485 (0.05380)
F-statistic	267	314
Adj. R-sqr	0.945	0.953
Observations	79	79

*Standard error in parenthesis. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$*

*Table 2, Fama and French model applied of the Agreement and Disagreement portfolio, 2013-2019*

We could argue why some of the control variables in the return regressions of Table 2 fail to show statistical significance, especially for the agreement portfolio. The lack of significance could be due to various reasons. For instance, the control variables are known predictors of returns, and there is evidence indicating lower post-publication return predictability for such predictors (McLean & Pontiff, 2016). Once an anomaly becomes public, investors may start trading to exploit it, which can destroy the uncovered pattern. Additionally, other studies have indicated a significant decline in return predictability after 2003 (Green, Hand, & Zhang, 2017). Similarly, Chordia, Subrahmanyam, and Tong (2014) demonstrated that capital market anomalies have weakened in recent periods (Chordia, Subrahmanyam, & Tong, 2014). Considering these studies, the insignificance of control variables is not so unexpected; the same problem was incurred by Brandon et al. (2021) in their study referred to ESG rating disagreement and stocks returns. However, the F-statistic indicates that we can confidently reject the null hypothesis which assumes that all control variables are equal to zero.

Let us now consider the stock analysis, carried to test whether there is a significant difference between the abnormal returns (alphas) of stocks with high and low ESG disagreement. In this case the stock selection is the one described in section 3.4., “Two-sample t-test”. Table 3 reports the t-test, the p-value and if the null can be rejected or not.

t-test	0.2309
p-value	0.8180
null hypothesis	not rejected

*Table 3, two sample t-test results, 2013-2019*

With a p-value of 0.8180 the null hypothesis can not be rejected, i.e., there is not a statistical difference between the average alphas of the two samples. In this case, stocks with high ESG disagreement do not present higher abnormal returns, in favour of the alternative hypothesis of the second research question of this paper.

Based on the graph of the cumulative returns of the two portfolios, with the agreement portfolio reaching a 208% raw return compared to 182% for the disagreement one, the evidence suggest a trade-off between the two strategies in favour of the agreement portfolio; however, from a statistical point of view, there is no difference in the abnormal returns at a stock level. Considered it all, the obtained results give enough evidence in favour of both the alternative hypotheses, namely, there not exists a performance trade-off by investing in a portfolio based on high ESG disagreement stocks and stocks with high ESG disagreement do not exhibit abnormal returns compared to the ones with low ESG disagreement.

## January 2013 – January 2023 sample

Similarly to the previous analysis, figure 3 reports the cumulative returns of the agreement and disagreement portfolios, in order to investigate the possible difference in choosing a strategy based on ESG disagreement. This time, the time period is of 10 years. Having one return for each month, the observations are 120.

Both strategies show a similar pattern, with an increasing trend over time especially after the beginning of 2020. Even though the correlation between the two time series is 0.996, the volume reached by the two strategies is quite different. The agreement portfolio, at the end of the sample, reaches a raw return of 343% compared to 260% for the disagreement one. The evidence suggests a performance trade-off in pursuing this strategy.

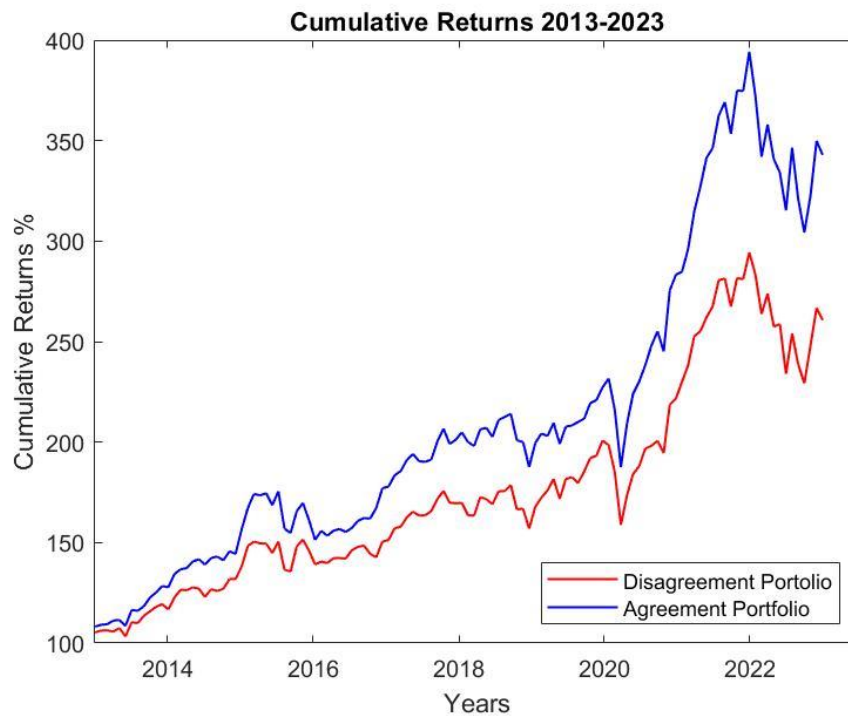


Figure 3, Cumulative Returns for the Agreement and Disagreement portfolio, 2013-2023

Table 4 reports the results of Fama and French model applied to the whole time period. The R-squared is still very high for both models, suggesting a good fit of the data. In this case both values of alpha are negative and non-significant, suggesting again no abnormal return associated with these strategies. The values of beta are significant in both cases at 1% level, slightly larger than one for the agreement portfolio, making it subject to a larger systematic risk.

	Agreement Portfolio	Disagreement Portfolio
	Coefficients	
Alpha	-0.00003 (0.00093)	-0.00140 (0.00085)
Rm - Rf	1.05690*** (0.02492)	0.99641*** (0.02267)
SMB	-0.03645 (0.07010)	-0.00448 (0.06377)
HML	-0.02807 (0.06559)	0.09075 (0.05966)
RMW	-0.01672 (0.05260)	-0.04380 (0.04784)
CMA	0.00211 (0.05262)	0.08088* (0.04786)
F-statistic p-value	474	540
Adj. R-sqr	0.952	0.958
Observations	120	120

Standard error in parenthesis. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

Table 4, Fama and French model applied of the Agreement and Disagreement portfolio, 2013-2023

The lack of statistical significance of the remaining control variables can be attributed to the same reasons reported before; again, the F-statistic is connected with a very low p-value. Even in this case the evidence is in favour of the alternative hypothesis of the first research question for the same reasons analysed in the previous section.

Regarding the stock analysis for the 2013-2023 time period, Table 5 reports the t-test, the p-value and if the null can be rejected or not.

t-test	-1.177
p-value	0.2416
null hypothesis	not rejected

Table 5, two sample t-test results, 2013-2023

Similarly to the previous analysis, with a p-value of 0.2416, the difference in the average values of alpha is not significantly different from zero. Again there is evidence in favour of the null hypothesis of the t-test and in favour of the alternative hypothesis of the second research question of this thesis.

Further analysis are presented in the appendix, such as the regression tables of the portfolio

analysis using robust standard errors and, separately, using Fama and French 3-factor model. All these variants lead to the same conclusions.

## 4.2. Discussions

The cumulative returns of the two time intervals considered impressively differ in the volumes reached. Right after 2019 the cumulative returns spiked marking a great difference between the two samples. The main driver of this difference rests in the volume of ESG investments that took over during 2019 and afterwards. In 2019 ESG funds' net money flows were almost four times as much as the \$5.5 billion in 2018, reaching \$20.6 billion in 2019 (Benjamin, 2021). This trend continued in 2020 when the global sustainable investment market was worth \$35.3 trillion, 16.67% more than 2019 (Global Sustainable Investment Review , 2020), and in 2021 and 2022.

At the same time, along with this increasing trend, the cumulative returns of the agreement portfolio is persistently above the disagreement one. According to research by Friede et al. (2015) and Whelan, Atz et al. (2020), there is a positive relationship between environmental, social, and governance (ESG) factors and financial performance (Friede, Busch, & Bassen, 2015). This theory is supported by several studies (see Whelan, Atz et al. (2020)), which suggest that companies with higher ESG ratings tend to perform better. According to these studies, if we consider the stocks with low ESG disagreement as those perceived as "truly" ESG by investors, then the agreement portfolio should generate higher cumulative returns, as it has been observed in practice. Moreover, according to the study conducted by Correia-Domingues et al. (2019), the level of sustainability is negatively related to the value at risk, supporting that higher scored funds offer better protection against extreme losses (Durán-Santomil, Otero-González, Correia-Domingues, & Reboredo, 2019). Therefore the disagreement portfolio, composed by stocks whose ESG rating is not commonly recognised among raters, should be more exposed to higher losses in volatile times.

Furthermore, the theoretical model does not report abnormal returns for any of the portfolio strategies and the agreement portfolio reports higher cumulative returns. Therefore, the null hypothesis of the first research question has to be rejected, i.e., there is no advantage in investing in a portfolio with high ESG disagreement.

Regarding the second research question, the stock level analysis does not provide evidence for higher (on average) abnormal returns related to stocks with high ESG disagreement leading to reject, even in this case, the null hypothesis.

The field of study for this thesis is quite recent and still unexplored, making this paper one of the few to investigate the relation between stock performance and rating disagreement. This implies that the previous literature on this topic is still developing and lacking at the same time, allowing the comparison of the obtained results with just a couple of other studies. Billio et al. (2020) prove that ESG disagreement disperses the effect of preferences of ESG investors on asset prices, to the point that even when there is agreement, the latter is so weak that it has no impact on the financial performances of ESG portfolios (Billio, Costola, Hristova, Latino, & Pelizzon, 2020).

On the other end, Brandon et al. (2021), find that stock returns are positively related to ESG rating disagreement and, in particular, to environmental rating disagreement, suggesting a risk premium for firms with higher ESG rating disagreement (Brandon, Krueger, & Schmidt, 2021).

According to the empirical analysis, the results are more in line with those of Billio et al., given no statistical difference in the average abnormal returns from the two sample t-test and given no significance in the values of alpha for the two portfolio strategies. As a consequence, both the null hypotheses of this paper are rejected.

### 4.3 Robustness analysis

In this section the robustness analysis is presented, aiming to strengthen the stability and reliability of this study. Both qualitative and quantitative aspects are considered.

#### 4.3.1 ESG conversion

The main discretionary decision made in this thesis that could bias the results is represented by the conversion method applied to Morningstar's data. The motivation for this conversion is to allow comparability between datasets, since the main variable of the study is represented by the ESG disagreement: to understand how much the ratings provided differ, they must be equally scaled, otherwise the comparison would not be meaningful. For this reason the data were rescaled on a 0-100 scale, the same adopted by Refinitiv. This robustness analysis aims to justify the choice made in order to strengthen and settle the reliability of this paper.

First, considering Morningstar's dataset, let us compare the average change in ratings each month, i.e. what is the average percentual change in the rating from one month to the next

one. This analysis will provide three average changes: one referred to the ratings from 2013 to September 2019, i.e., to the original data before the conversion; the second one is referred to the percentual change in ratings after September 2019, but the change is calculated on the original data, not converted in the 0-100 scale; the last one represents the percentual change related to the converted dataset, i.e., of the converted ratings between September 2019 and the end of the sample.

In addition, the overall change is reported. It represents the mean between the average change before September 2019 (*original dataset before 09/2019* in the table) and, respectively, the average change after September 2019 of the original (*original dataset after 09/2019*) and the converted dataset (*converted dataset*).

	Average % change	Overall change
Original dataset before 09/2019	0.90%	0.60%
Original dataset after 09/2019	0.30%	
Converted dataset	0.50%	0.70%

*Table 6, comparison between the change in rating between the original dataset and the converted one*

As table 6 reports, both the original and converted dataset show less % changes after the conversion, from 0.9% to 0.3% and 0.5%. The converted sample only shows a 0.2% increase in change with respect to the original sample; this implies that the conversion maintains, on average, the changing tendency of the original dataset. Moreover, the average overall change throughout the whole sample is nearly identical, as the last columns reports.

### 4.3.2 ESG rating sources

One of the key distinctions between this study and prior research lies in the source of the ESG ratings utilized. As a graduating student, my access to datasets is restricted and contingent on the University's arrangements with data suppliers. Consequently, I had access to only two datasets, namely Refinitiv and Morningstar, which compelled me to create the ESG disagreement metric solely from these sources.

This is an important difference from prior research, which typically incorporates four or more sources to establish a more robust and reliable measure of disagreement. The limited access to datasets constrained my ability to gather a diverse range of ESG ratings sources,



which could have potentially impacted the reliability and comprehensiveness of the analysis. Despite these limitations, the research project still offers valuable insights into the ESG ratings disagreement among various firms and their implications for investment decisions. The findings of this study should be interpreted with the understanding that the dataset utilized is more limited in scope than those of previous research projects, and that future studies that incorporate a broader range of ESG rating sources may yield different results.

### 4.3.3. Portfolio selection

The selection of the stocks strictly depends on which percentiles are considered as thresholds. For instance, choosing the 30<sup>th</sup> and 70<sup>th</sup> percentiles will include more stocks in the analysis, including even those stocks whose disagreement is lower. One could argue that selecting the 10<sup>th</sup> and 90<sup>th</sup> percentiles would be a more accurate approach, ensuring that the considered stocks truly reflect the level of agreement and disagreement among rating agencies. To test the robustness of the results reported in this stock analysis, the two sample t-test is implemented again using the 10<sup>th</sup> and 90<sup>th</sup> percentile as threshold. The table below displays the outcomes of this analysis.

	2013-2019	2013-2023
t-test	0.1451	-0.6143
p-value	0.8850	0.5402
null hypothesis	not rejected	not rejected

*Table 7, two sample t-test results for both considered periods, with 10<sup>th</sup> and 90<sup>th</sup> percentile portfolios*

The p-values for the two time periods are both above 5%, providing evidence in favour of the null hypothesis.

These results strengthen the main findings, demonstrating that they remain robust even when a narrower pool of stocks is chosen.

## 5. CONCLUSIONS

With the ESG world increasing in size and importance, the availability of reliable sources for ESG ratings becomes essential. This need shed light on the issue of ESG rating disagreement, i.e., the fact that different rating agencies have varying criteria and methodologies for evaluating a company's ESG performance. This has led to discrepancies in ratings, making it difficult for stakeholders to determine which companies can be considered truly sustainable. The lack of consensus on ESG ratings is not only a challenge for investors but also for companies that are trying to improve their ESG performance. Companies may receive different ratings from different agencies, which can create confusion among stakeholders and make it challenging to prioritize sustainable initiatives.

While several studies focus on the reasons behind the disagreement, this paper provides further knowledge on the consequences of the disagreement by studying its relationship with stock and portfolio returns. According to the cumulative returns, the agreement portfolio shows a better performance compared to the disagreement one, especially after the pandemic, indicating that there is a performance trade-off in choosing the agreement portfolio. Moreover, statistical evidence at stock level indicates that there is no difference between the average abnormal returns of the strategies based on ESG disagreement.

The contribution of this paper rests in its important practical consequences. First of all, these results are crucial for investors who are looking for sustainable investment strategies. According to Brandon et al. (2021), asset managers and investment managers in the US who wish to maximize financial performance while investing sustainably, should take into account the effect of rating disagreement on stock returns (Brandon, Krueger, & Schmidt, 2021). In this case, however, the evidence points in a different direction, more in line with Billio et al. (2020), suggesting no effects due to the disagreement.

Secondly, this study has implications on how companies will pursue their sustainable goals in the future and, at the same time, it should help to disincentive greenwashing. Let us consider again the results obtained by Brandon et al. (2021), in which investors perceive a dispersed ESG performance as an additional source of risk, requiring a separate risk premium. In this context firms could be incentivized to report ambiguous ESG performances in order to be rated differently by the rating agencies and to see their stocks selected by investment funds. So, paradoxically, there could be advantages in greenwashing. On the

contrary, according to this study there is no advantage in having rating discrepancy, reducing any possible incentive in pursuing greenwashing strategies.

To conclude, this study makes a valuable contribution to the field of sustainability. While the relationship between ESG ratings and stock returns has been extensively examined, there has been limited research on the impact of rating discrepancies on stock returns. This paper is the first to investigate this issue in Europe, adding to the existing body of knowledge. Additionally, previous studies have yielded conflicting results, indicating that the debate is ongoing and any new insights can be valuable.

### 5.1. Further research

This study serves as a solid foundation for future research. An interesting study could involve the same analysis presented in this paper, but retrieving data from more rating agencies. The agreement and disagreement portfolio could be differently composed and the results could point in a different direction. In that case, the disagreement measure can be represented by the standard deviation as in Brandon et al. (2021), increasing the comparability of the results. The availability of more datasets could even involve a study focused on the effect of each component (E, S, G) on stock returns.

Moreover, the same study can be conducted in some years from now. In recent years the concern for ESG disagreement dramatically increased; in 2022 at least four regulators - including the European Securities and Markets Authority and the Securities and Exchange Board of India - have proposed regulating the ESG ratings sector (Illago, 2022). In addition, several regulations were disclosed such as the SFDR (Sustainable Finance Disclosure Regulation, 2021) and the EU Taxonomy Regulation (Regulation EU, 2020/852). It all suggests that in the near future the disagreement could be partially reduced, yielding different results from the study.

Furthermore, this study can be replicated in different markets and countries, to test whether the results would be the same. Perhaps the disagreement is perceived differently in different countries leading to different implications.

Anyway, the lack of studies regarding ESG disagreement and the lack of consensus on its effects make it a perfect topic to delve into.



## References

- (2004). Retrieved from Who Cares Wins:  
[https://www.unepfi.org/fileadmin/events/2004/stocks/who\\_cares\\_wins\\_global\\_compact\\_2004.pdf](https://www.unepfi.org/fileadmin/events/2004/stocks/who_cares_wins_global_compact_2004.pdf)
- (2020, January 3). Retrieved from Rinnovabili.it:  
<https://www.rinnovabili.it/ambiente/greenwashing-shell-promette-ma-non-mantiene/>
- Avramov, D., Cheng, S., Lioui, A., & Tarelli, A. (2022). Sustainable Investing with ESG rating uncertainty . *Journal of Financial Economics*, 642-664.
- Benjamin, J. (2021, January 14). *ESG funds ride 'perfect storm' to record inflows*. Retrieved from InvestmentNews: <https://www.investmentnews.com/esg-funds-ride-perfect-storm-record-inflows-176396>
- Berg, F., F. Kolbel, J., & Rigobon, R. (2022). Aggregate Confusion: The Divergence of ESG ratings. *Review of Finance, Oxford University*, 2-3.
- Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2020). Inside the ESG Ratings: (Dis)agreement and Performance. *SAFE working paper*, 42.
- Bowen, H. (1953). *Social responsibilities of the businessman*.
- Brandon, R. G., Krueger, P., & Schmidt, P. S. (2021). ESG rating disagreement and stock returns. *Financial Analysts Journal*.
- Carroll, A. (1999). Corporate Social Responsibility, Evolution of a Definitional Construct. *BUSINESS & SOCIETY*, 3.
- Carroll, A. B. (1991). The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders. *Business Horizons*, 39-48.
- Chordia, T., Subrahmanyam, A., & Tong, Q. (2014). Have Capital Market Anomalies Attenuated in the Recent Era of High Liquidity and Trading Activity? *Journal of Accounting and Economics*.
- Davis, K. (1960). Can business afford to ignore social responsibilities? *California Management Review* , 70-76.
- Durán-Santomil, P., Otero-González, L., Correia-Domingues, R. H., & Reboredo, J. C. (2019). Does Sustainability Score Impact Mutual Fund Performance? . *Sustainability*.
- Eccles, R. G., & Klimenko, S. (2019). The investor revolution. *Harvard Business Review*.
- Fama, E. F., & French, K. R. (2014). A five-factor asset pricing model. *Journal of Financial Economics*.
- French, K. R. (2023). *Data Library*. Retrieved from Kennet R. French:  
[https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)

- Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 210-233.
- Friedman, M. (1970). The social responsibility of business is to increase its profits. *New York Times Magazine*.
- Funds-Europe. (2023, June 27). *European ESG assets to reach €9 trillion by 2025*. Retrieved from Funds Europe: <https://www.funds-europe.com/news/european-esg-assets-to-reach-9-trillion-by-2025>
- Global Sustainable Investment Review*. (2020). Retrieved from Global Sustainable Investment Alliance: <https://www.gsi-alliance.org/wp-content/uploads/2021/08/GSIR-20201.pdf>
- Global Witness*. (2023, February 1). Retrieved from Global Witness: [https://www.globalwitness.org/en/campaigns/fossil-gas/shell-faces-groundbreaking-complaint-misleading-us-authorities-and-investors-its-energy-transition-efforts/?utm\\_source=hootsuite&utm\\_medium=twitter\\_](https://www.globalwitness.org/en/campaigns/fossil-gas/shell-faces-groundbreaking-complaint-misleading-us-authorities-and-investors-its-energy-transition-efforts/?utm_source=hootsuite&utm_medium=twitter_)
- Green, J., Hand, J. R., & Zhang, X. F. (2017). The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *The Review of Financial Studies*.
- Gunnar, F., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 21-233.
- Illago, H. (2022, October 14). *ESG ratings space needs regulatory intervention*. Retrieved from Institute for Energy Economics and Financial Analysis: <https://ieefa.org/resources/esg-ratings-space-needs-regulatory-intervention>
- McLean, R. D., & Pontiff, J. (2016). Does Academic Research Destroy Stock Return Predictability? *Journal of Finance*, 5-32.
- Monasterolo, I., & De Angelis, L. (2020). Blind to carbon risk? An analysis of stock market reaction to the Paris Agreement. *Ecological Economics*, 170.
- Morningstar*. (2021). Retrieved from <https://connect.sustainalytics.com/hubfs/SFS/Sustainalytics%20ESG%20Risk%20Rating%20-%20FAQs%20for%20Corporations.pdf>
- PRI*. (2021). Retrieved from <https://www.unpri.org/annual-report-2021>
- PwC. (2022, October 10). *PwC*. Retrieved from <https://www.pwc.com/gx/en/news-room/press-releases/2022/awm-revolution-2022-report.html#:~:text=ESG%2Dfocused%20institutional%20investment%20seen,assets%20under%20management%3A%20PwC%20report>
- Refinitiv*. (2022). Retrieved from [https://www.refinitiv.com/content/dam/marketing/en\\_us/documents/methodology/refinitiv-esg-scores-methodology.pdf](https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf)
- Stobierski, T. (2021, June 15). *EYE-OPENING CORPORATE SOCIAL RESPONSIBILITY STATISTICS*. Retrieved from Harvard Business School: <https://online.hbs.edu/blog/post/corporate-social-responsibility->



## Appendix

The table below reports some examples of how the rating conversion is applied. The first two columns report the name of the stock with the corresponding ticker, while the remaining columns report the rating right before the conversion, the original rating after the conversion and the converted rate. The last column reports the difference between the rate before the conversion and the converted one.

Stock	Rating August 2019	Rating September 2019	Rating September 2019 converted	Difference Rating August and September Converted
Covestro AG	79	9.995	80.01	1.01
American Airlines	48	23.585	52.83	4.83
Actividades de Construcción	74	12.59	74.82	0.82
Abold Delhaize	50	23.095	53.81	3.81
Adidas AG	68	15.69	68.62	0.62
Evotec SE	43	26.9	46.2	3.2
Ashford Hospitality	76	11.44	77.12	1.12
Alfa Financial Software	66	16.6	66.8	0.8
CaixaBank	74	12.895	74.21	0.21
Andrea Electronics	77	11.74	76.52	-0.48
Acciona Energias	54	20.07	59.86	5.86
Amofagusta	64	20.29	59.42	-4.58
ARCADIS	85	10.95	78.1	-6.9
Argenx SE	76	13.34	73.32	-2.68

Table 8, examples of rating conversion



Following, the graphs of the cumulative returns of each factor created for the Fama-French model. Only the time period 2013-2023 is reported, since it already incorporates the graphs referred to the period 2013-2019.

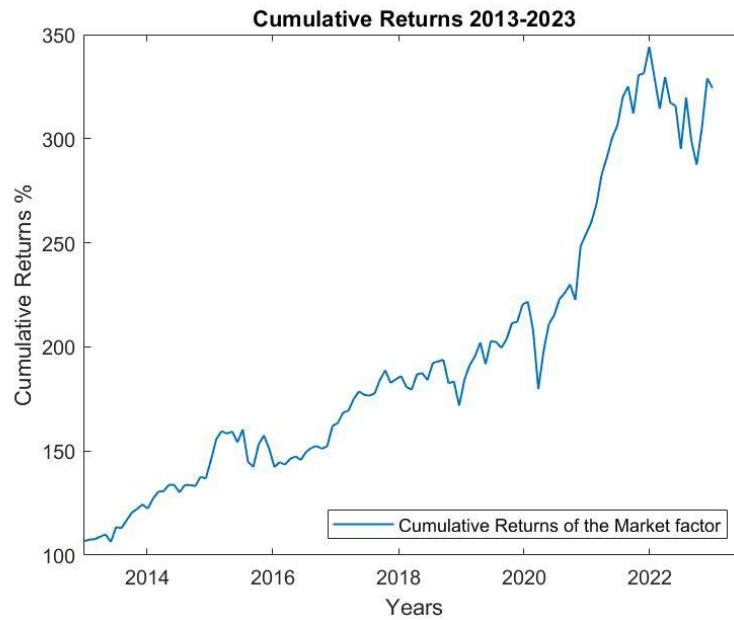


Figure 4, Cumulative Returns of Market factor

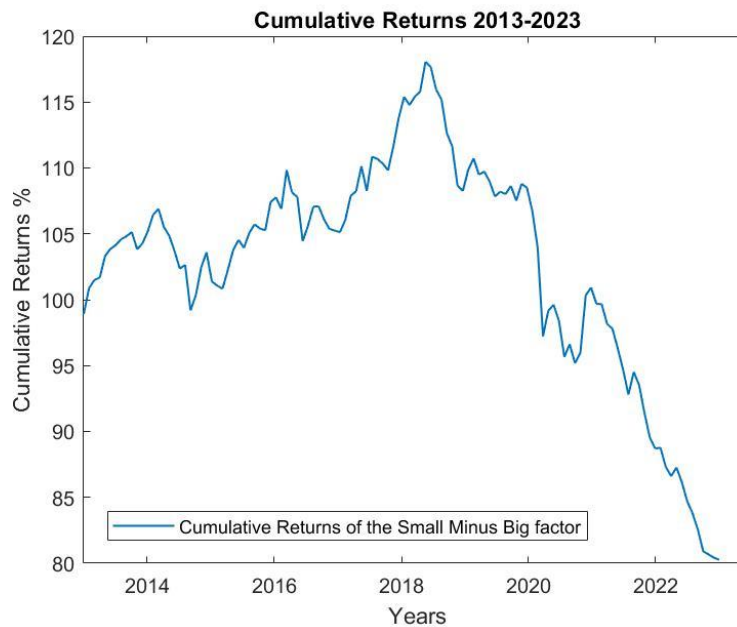


Figure 5, Cumulative Returns of Small Minus Big

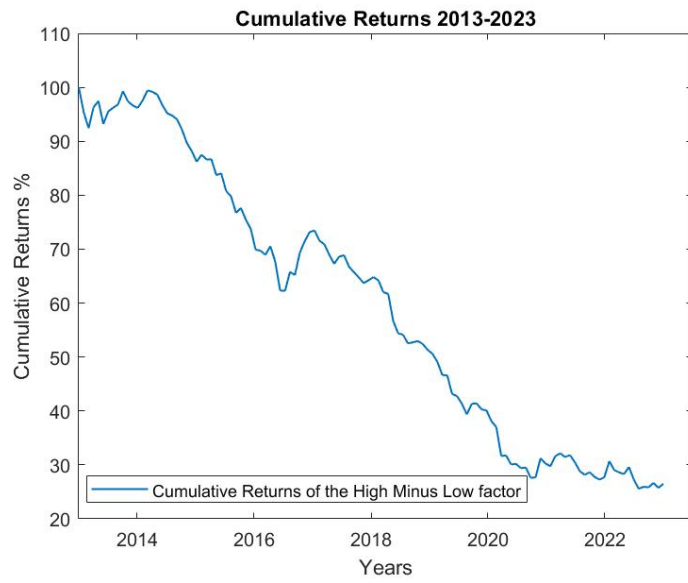


Figure 6, Cumulative Returns of High Minus Low factor

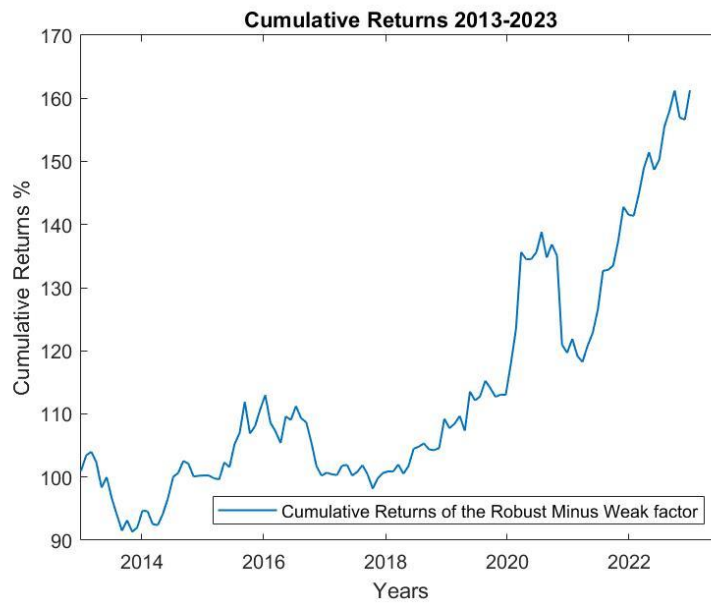


Figure 7, Cumulative Returns of Robust Minus Weak factor

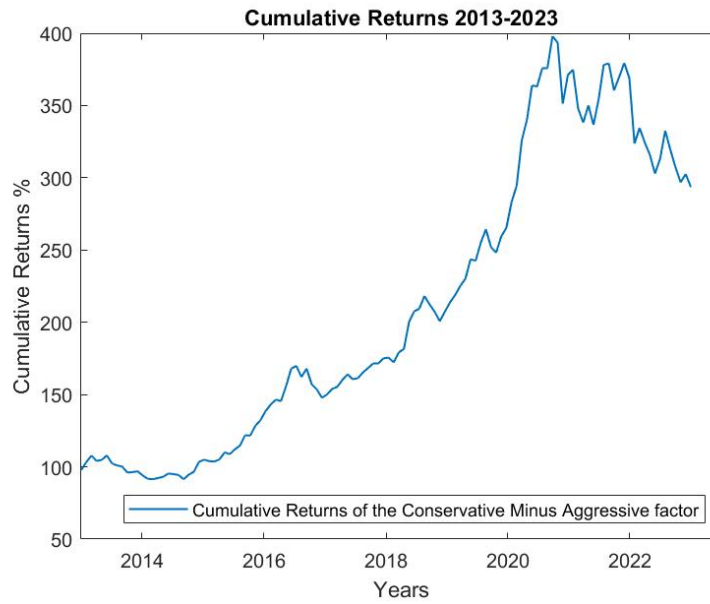


Figure 8. Cumulative Returns of Conservative Minus Aggressive factor

The tables below report the results of Fama and French 3 factor model applied to the portfolios. The results are in line with those reported in the main study.

	Agreement Portfolio	Disagreement Portfolio
	Coefficients	Coefficients
Alpha	0.00107 (0.00104)	-0.00216** (0.00095)
Rm - Rf	1.00460*** (0.02902)	0.99864*** (0.02652)
SMB	-0.09394 (0.07331)	0.11622* (0.06700)
HML	0.05519 (0.03689)	-0.06745** (0.03372)
F-statistic	454	503
Adj. R-sqr	0.946	0.951
Observations	79	79

Standard error in parenthesis. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

Table 9, Fama and French 3 factor model applied of the Agreement and Disagreement portfolio, 2013-2019

	Agreement Portfolio	Disagreement Portfolio
	Coefficients	Coefficients
Alpha	-4.885e-05 (0.00092)	-0.00152* (0.00085)
Rm - Rf	1.0599 *** (0.02260)	1.0116*** (0.02088)
SMB	-0.03003 (0.06192)	0.05159 (0.05721)
HML	-0.02419 (0.02650)	0.02018 (0.02448)
F-statistic	802	886
Adj. R-sqr	0.953	0.957
Observations	120	120

Standard error in parenthesis. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

Table 10, Fama and French 3 factor model applied of the Agreement and Disagreement portfolio, 2013-2023

The tables below report the results of Fama and French 5 factor model when robust standard errors are considered. The results are in line with those reported in the main study.

	Agreement Portfolio	Disagreement Portfolio
	Coefficients	Coefficients
Alpha	-0.00102 (0.00100)	-0.00250** (0.00094)
Rm - Rf	1.0225*** (0.02982)	0.96891*** (0.02785)
SMB	-0.07995 (0.07791)	0.11261 (0.07274)
HML	-0.01301 (0.07558)	-0.09856 (0.05966)
RMW	0.05375 (0.05669)	-0.12713** (0.05293)
CMA	-0.052343 (0.05795)	0.02511 (0.05411)
F-statistic	291	306
Adj. R-sqr	0.949	0.951
Observations	79	79

Robust standard error in parenthesis. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

Table 11, Fama and French 5 factor model, with robust standard errors, applied of the Agreement and Disagreement portfolio, 2013-2019

	Agreement Portfolio	Disagreement Portfolio
	Coefficients	Coefficients
Alpha	-0.00002 (0.00087)	-0.00080 (0.00084)
Rm - Rf	1.06210*** (0.02332)	0.97324*** (0.02248)
SMB	-0.04011 (0.65590)	-0.01650 (0.06322)
HML	-0.08349 (0.06137)	0.03995 (0.05915)
RMW	-0.00939 (0.04921)	-0.09019* (0.04744)
CMA	-0.03461 (0.04923)	0.05001 (0.04745)
F-statistic	474	529
Adj. R-sqr	0.952	0.957
Observations	120	120

*Robust standard error in parenthesis. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$*

*Table 12, Fama and French 5 factor model, with robust standard errors, applied of the Agreement and Disagreement portfolio, 2013-2023*