LUISS GUIDO CARLI

Department of Economics and Finance

Master's Degree in Finance

Do European Stock Prices react to ESG rating announcements?

SUPERVISOR

P. Santucci De Magistris

CANDIDATE

Silvio Petruzzi

Academic Year 2023/2024

Abstract

This study examines the short-term market reaction to Environmental, Social, and Governance (ESG) rating announcements for European companies from 2016 to 2023. Using an event study methodology with multiple event windows and three returns estimation models, both aggregate ESG ratings and single pillar ones are analyzed at first, then only the combined ESG rating is scrutinized for relevant sector-based sub-samples, specifically focusing on the energy, banking, and utility industries. Most of the observed Cumulative Average Abnormal Returns (CAARs) are positive but lack statistical significance in the aggregate analysis. Notably, downgrades show consistently higher CAARs than upgrades, with only a significant positive reaction to Environmental pillar downgrades in the five-day event window. The sector-specific analysis yields more pronounced results. Positive and significant CAARs across most of the event windows are observed after a decrease in the combined ESG rating in every sector considered. Conversely, rating upgrades cause negative CAARs in the energy sector, positive ones for utility stocks, and no particular reaction for banks. These findings suggest that while on average ESG information may not be immediately value-relevant to investors, except maybe for specific factors like the Environmental one, it can cause substantial unexpected price movements in specific sectors. The results of this study hope to contribute to the understanding of the integration of ESG factors in the investment decision process.

Contents

1 Introduction **3**

2 Literature review **7**

3 Methodology **11**

3.1 Event study methodology 11

3.2 Tests 19

3.3 Definition of event and estimation windows 19

4 Empirical Analysis **22**

4.1 Data 22

4.2 Summary statistics 25

4.3 Results 28

5 Conclusions **38**

Chapter 1

Introduction

The analysis carried on in this study is aimed at finding an answer to the question ”Do European stock prices react to ESG rating announcements?”, but where does this question come from and why it may be worth an answer?

In the past decade, much research has focused on the relationship between Environmental, Social and Governance (ESG) factors and corporate financial performance. Substantial evidence has been collected about the correlation between these metrics. A recent meta-study from NYU Stern Center for Sustainable Business and Rockefeller Asset Management (Whelan et al., 2021) examined more than 1,000 individual papers published in the period 2015-2020 and drew, among others, the following conclusions: there is a positive relationship between ESG and financial performance, and this becomes more marked over longer horizons; ESG investing provides downside protection during social or economic crises, and ”Sustainability initiatives at corporations appear to drive financial performance due to factors such as improved risk management and more innovation”.

These findings suggest that ESG variables can convey relevant information on a firm’s future performance and can be important signals to consider when making investment decisions. Indeed, while some argue that there is no novelty in the publication of these variables because most of the data on which they are constructed are publicly available, as well as information about what the firm does, it is worth noticing that an ESG rating is not only the result of summation of ESG news. New information is created by finding signals in unstructured data using complex models not available to everyone. Moreover, there is a valuation about the materiality of the news that cannot easily be done by a normal investor and agencies have actually also access to non-public information about the activity of a rating target, which is given to them throughout the rating definition procedure (Glück et al., 2021). Indeed, empirical studies demonstrate that investors are already factoring these ratings into their decision-making processes (Rzeźnik et al., 2022)(Leite and Uysal, 2023).

ESG ratings, thus, are variables loaded with information, they have predictive power over the future financial performance of a firm and are already capable of influencing investor behavior. Nevertheless, the importance of these ratings may be threatened by the great divergence across those offered by different providers (Berg et al., 2022), and this weakness has always to be considered when talking about ESG matters. The average correlation between ratings of different agencies is only 0.54, and ranges from 0.38 to 0.71. This heterogeneity can be explained by the subjectivity that still characterizes the evaluation process of ESG scores, with measurement (how a variable that is relevant to the assessment of each score is measured) being the main cause of it. Thus, while an agency may do great work summarizing an enormous amount of data in a single letter or number, the investor that receives this valuable information, may not be able to distinguish it from the noise, severely reducing its usefulness.

Despite this, numerous findings suggest that ESG information is important and relevant for investors. It seems thus worth inspecting how the market reacts to the release of news that fall in this category to see what impact do they have in the short term on stocks' returns and how quickly they are incorporated into prices. While rare studies have examined the immediate market reaction to ESG rating changes in the American context (Glück et al., 2021), the European market seems to be comparatively even less covered. This gap is particularly notable given Europe's pioneering role in ESG initiatives and regulation.

Using an event study methodology, this research examines both aggregate ESG ratings and Individual Environmental, Social, and Governance pillar scores for a comprehensive sample of

European companies. The analysis utilizes data from the Morgan Stanley Capital International (MSCI) Intangible Value Assets database, covering firms across all 27 European Union member states and the United Kingdom from May 2016 to August 2023. The market reaction is analyzed through multiple event windows, to account for the nature of the information, which, given its ”softness”, may take longer to become input of investment decisions. While, to ensure robustness, the study employs three models for estimating normal returns: a single-factor model, a three and a five-factors one. Additionally, the analysis is performed on three sector-specific sub-samples of stocks - Energy, Banking, and Utility - considering only changes in the combined ESG score as events. This approach aims to identify interesting patterns that may be averaged out in the original, broader sample.

Predominantly positive Cumulative Averaged Abnormal Returns were observed across the various scenarios, but most lack statistical significance. Notably, CAARs are consistently higher when they follow a rating decrease, particularly if it affects a single pillar. The most significant finding is a statistically significant positive market reaction following downgrades in the Environmental pillar, observed in the short-term event window, which includes the day of the event and the five trading days after it, that fades when wider windows are considered. This reaction is consistent across all three estimation models employed and is evidence of the efficiency of the market to incorporate rapidly news into prices.

Overall, the results of the first analysis suggest that, while there are observable trends in the market as a response to ESG rating changes, they are not consistent enough to rule out the hypothesis that they are a result of random variation. Therefore, ESG information may not be as immediately value-relevant to investors as other types of financial information, at least in the short term.

The analysis of the three sub-samples reveals interesting results that did not emerge in the aggregate one. The variation of the ESG combined score, which at first did not lead to any significant reaction, generates statistically significant CAARs that differ in size and direction across the three sectors. Upgrades of the ESG variable generate unexpected negative returns for stocks in the energy sector, unexpected positive returns for those in the utility sector, and no particular reaction in the banking sector. Downgrades instead, cause positive CAARs of different sizes in all three sectors. All findings are robust across the three models employed. Interestingly, CAARs from the sub-sample analyses are statistically different from zero also for wider event windows, while on aggregate only the [0,5] window showed significant results. This may suggest that prices react at different velocities to this kind of news depending on the sector.

This study aims to contribute to the growing literature on the impact of ESG in finance by focusing on the reaction of stock prices to changes in ESG ratings of European firms instead of the more covered American ones. While the scope and the implication of this research may be limited, the findings may still provide useful insights for future research in this area.

The thesis is structured as follows: the first two chapters define the research question and review the existing literature on the relationship between ESG and the financial market. They also underline the skewness of the research towards the American market and the long-term impacts of ESG news over short-term ones. The third chapter better defines the goal of the analysis and illustrates the methodologies used to reach it. The fourth one presents the data on which the analyses are performed and exposes and explains the results. The last chapter is a summary of what has been done.

Chapter 2

Literature review

This thesis aims to investigate short-term market reactions to ESG rating announcements, using the well-established event study methodology, which, since the 70s with the publication of a famous paper by Fama (1969), has become widely used to test market efficiency to adjust to new information. The premise of this tool is that, in an efficient market, stock prices should shortly deviate from their expected trajectory if new information, relevant to investors, is made public. Two early papers from Brown and Warner (1980, 1985) discuss a variety of issues that arise when performing such analysis, from the choice of the model used to estimate expected security price performances to the use of daily instead of monthly returns. Later, MacKinlay (1997) gives a precise outline of how to design an event study, while, more recently, Kothari and Warner (2004) gives an overview of how this methodology changed over time, underlying the utility of the tool in testing short-term market efficiency and expressing its concerns about the efficacy of long-horizon event studies. This thesis aims to investigate how the European stock market reacts to announcements of changes in ESG variables. If they hold importance to investors, abnormal returns - returns that deviate from the normal, expected ones - should be observed in the days immediately following the announcement of the change. But why investors should care about these variables?

The relationship between ESG and financial performance has been extensively scrutinized since the 1970s, when the concept of socially responsible investing emerged. In a famous early paper Eccles et al. (2014) analyzes the performance of different firms, grouped as high and low sustainability ones, over a period of 18 years, and finds that firms in the first group outperformed both in stock market and accounting terms the ones in the second. While he refrains from claiming a causal relationship between the variables he highlights the cultural significance of the findings, which suggest that firms can 'do well by doing good'. Papers written on the subject are numerous and often have contradictory results, for this reason, it is very valuable the work made by Gunnar Friede and Bassen (2015). It is focused on what more than 2,200 academic studies written from the 1970s to 2014 have in common, rather than the disparities among them, and shows that the large majority (close to 90%) of them exhibit positive findings on the relationship between ESG and corporate financial performance, stable over time and markets. More recent research continues to support these findings. Another meta-study, previously mentioned in the introduction (Whelan et al., 2021), examines 1,000 papers published between 2015 and 2020 on the subject. It reinforces the notion of a positive link between ESG and financial performance, which seems to be particularly marked over extended time frames. The study also underlines the capacity of ESG investments to provide downside protection during turbulent times and to push corporations toward better risk management and innovation. Further research suggests that robust ESG practices may be capable of mitigating risks, leading to better financial outcomes during social and economic crises, such as the COVID-19 pandemic (Broadstock et al., 2021). There is evidence of this risk-reducing force for European enterprises as well (Sassen et al., 2016). Moreover, superior performance in ESG is also linked to easier access to capital. Firms that allocate time and resources to corporate social responsibility (CSR) activities are less likely to incur short-term opportunistic behavior and more likely to be transparent to the market about their activities, trying to signal their long-term focus and differentiate themselves from peers (Cheng et al., 2013)(Chen et al., 2023). To quantify this effect, Li et al. (2024) found that firms subject to ESG ratings face a 0.46% lower cost of capital in the Chinese market compared to firms that do not.

It seems thus safe to assert that the ESG and financial performance of a firm are strictly intertwined. Consequently, investors are starting to use ESG variables as inputs in their financial decisions. Hartzmark and Sussman (2019) found evidence that mutual fund investors are not indifferent to ESG ratings. They treat sustainability as a positive fund attribute, allocating more money to funds rated positively than to those rated poorly, and do this by showing a somehow irrational behavior. In fact, there are no significant differences between inflows and outflows of funds rated with intermediate scores, but only between the ones that have the best and the worst ratings. This may suggest that, in this field, investors are prone to react only to extreme outcomes of discrete changes, without questioning too much what underlies them. Further evidence of a particular investor reaction to rating changes is found by Rzeźnik et al. (2022), who noticed that investors trade based on a rating change without questioning it. Their study exploited an inversion in the Sustainalytics rating methodology, after which better ESG scores were represented by lower numbers and not the opposite, as it is in most cases, to show that investors continued to use the change of rating as a signal but started to trade in the wrong direction because they misinterpreted the news. Also Leite and Uysal (2023) found evidence of the fact that ESG ratings significantly influence how investors react to new information about firms, giving less weight to negative news regarding the pecuniary performance of the ones with higher ESG scores.

Despite the significant body of evidence linking ESG performance to long-term financial returns, as well as research indicating that investors are influenced by ESG metrics in their decisions, there is a notable gap in the literature when it comes to understanding how markets react in the short-term to the announcement of changes in these variables. Most of the research focuses on the U.S. market, with only a limited number of studies examining the European one. This makes the gap even more interesting given the centrality of Europe in the ESG landscape. The main findings are the following: Serafeim (2023) found that consensus ESG ratings predict future ESG news and that positive, financially material ESG news is followed by more positive industry-adjusted returns while negative news by more negative ones. Furthermore, he outlined that the positive reaction is smaller for firms with higher ESG scores and that the market reaction, regardless of the direction, is magnified if there is agreement across rating providers.

An analysis that is more similar to the one carried on in this thesis, despite being still focused on the American market, is the one from Glück et al. (2021) which is an event study examining the effect of ESG rating changes on stock prices. They found significant negative cumulative averaged abnormal returns within eleven days following the event of a downgrade in the environmental (E) or social score (S), while they did not find statistically significant effects for upgrades. They underline this asymmetry across results, attributing it to the fact that while a downgrade in an ESG score compels exclusion from the portfolio of investors with certain guidelines to follow, an upgrade merely broadens the investible universe of stocks from which they can choose. Another plausible explanation (Ederington and Goh, 1998) is that firms may disseminate positive information in advance while being quite hesitant to reveal negative ones, leading to negative abnormal returns the day on which they are released.

Lastly, it is worth noticing that the market reaction that follows an increase or a decrease in an aggregated ESG or pillar rating is not obvious. While the cited literature about the relationship between ESG and long-term financial performance may drive some to expect an increase in the stock price after the announcement of a higher rating, findings about the short-term effect of this kind of news are sparse and diverse. The effect that ESG performance can have in the short term is ambiguous, and some reasons may justify a movement in prices opposite to that of the ratings. Serafeim (2023) underlines the possibility that the cost of the investment needed to achieve a rating increase may outweigh the potential gains of a better firm image and that managers may be willing to invest in negative NPV projects to increase their ”green” reputation. Another example of what may cause an ambiguous market response is a decrease in the social pillar rating, which signals a poor social performance from the community perspective, caused by a reduction in the workforce. The firm’s decision could lower the expenses and bring to higher profitability in the future, meaning that the announcement of the rating decrease may cause a positive price reaction.

The fuzziness of the findings about the short-term market reaction and the uncertainty about what to expect from them make the analysis even more interesting, considering also the lack of relevant studies with a focus on the short-term reaction of the European market, despite it being one, if not the most relevant one when considering ESG issues.

Chapter 3

Methodology

3.1 Event study methodology

Even though the first event studies date back to the 1960s, when numerous scholars were interested in testing the newly formed efficient market hypothesis, their basic statistical format has not changed much over time (Kothari and Warner, 2004) and remains similar to the one outlined in the stock split event study by Fama (1969). Since then, the event study methodology has become an important part of financial economics, but only two main changes in methodology have taken place: the use of daily rather than monthly observations, and the implementation of slightly more sophisticated models to estimate abnormal returns.

Two early papers from Brown and Warner (1980, 1985) and a more recent paper from MacKinlay (1997) give a broad outline of this tool. Another, more recent paper from Kothari and Warner (2004), also has been a useful survey.

The event study methodology has become so widely used for its simplicity and its usefulness in providing arguments to important and vastly debated topics, such as the one on the above-mentioned market efficiency. Systematically nonzero abnormal security returns that persist after a particular corporate event are, indeed, inconsistent with the Efficient Market Hypothesis (EMH), in particular with its semi-strong formulation, which states that all publicly available information is quickly and fully reflected in prices (Brown and Warner, 1980). If the market is efficient, then, new information should almost instantaneously be incorporated into prices, and there should not be significant abnormal returns days after the announcement date.

In the corporate context, the event study methodology provides a measure of how and how much an unanticipated event impacts the firm's value, while "event studies focusing on announcement effects for a short-horizon around an event provide evidence relevant for understanding corporate policy decisions" (Kothari and Warner, 2004). Event studies are also useful outside of the financial world. They are vastly utilized in accounting to measure the impact of earnings announcements on stock prices, and in law and economics, where they are useful to quantify the effect of a regulation or of damages in legal liability cases (Kothari and Warner, 2004).

In this study, the methodology is used to assess whether ESG information is relevant for investors in the short term. The literature cited in chapter two seems to suggest that ESG ratings bring relevant information about future firms' performance and that investors are influenced in their decisions by these variables. It is thus worth questioning whether announcements of changes in ESG ratings are considered sufficiently important to cause unexpected abnormal returns immediately after their publication. Such a finding would stress the significance of these variables for the European market and highlight the role of rating agencies in delivering them.

To test whether an event has produced returns that were not expected by the market it is at first necessary to model the ones that could have been experienced had the event not taken place. These returns are called "normal returns" and can be estimated by a variety of models. At first, the more widely used ones were the market model and the constant mean return model, two estimation methods that are simple and provide solid results. The first, represented in equation 3.1, is a model that estimates the normal return R_{it} for stock i at time t as the mean return of security i (μ_i) over a certain window, plus a disturbance term ζ_{it} with mean 0 and variance $\sigma^2_{\zeta_i}$ (MacKinlay, 1997):

$$R_{it} = \mu_i + \zeta_{it}. \tag{3.1}$$

The second is a statistical model that is very similar to the CAPM but does not subtract the risk-free rate to any of the two sides of the equation.

The effort of collecting and cleaning extra data as inputs for more complex models, may not always be worth it, as outlined by Armitage (1995) who wrote that ”beyond a simple, one-factor market model, there is no evidence that more complicated methodologies convey any benefit”. This idea has been strengthened by various authors. MacKinlay (1997) states that generally, gains from employing multifactor models are limited, because the marginal explanatory power of additional factors is small and the consequent reduction in the variance of the abnormal returns, which are the error from the regression, is little. Brown and Warner (1980) even state that the mean return model, despite its simplicity, often yields results that are similar to those of more complicated models (also if this statement probably better relates to the use of monthly returns).

Anyway, given the easier accessibility of data today compared to 1994, other slightly more sophisticated models have become widely used in the event study literature (Kothari and Warner, 2004), the Fama-French factor models. These models extend the traditional Capital Asset Pricing Model (CAPM), where the market risk premium is the sole driver of extra returns, by incorporating additional risk factors. At first, they discovered the size and the value factors (Fama and French, 1993). Then, in 2013, the two professors actually extended the model incorporating two extra risk factors (Fama and French, 2014). In this analysis, normal returns are estimated for robustness in three different ways: with a normal CAPM and with the two extension of it just discussed. The five key factors included in the latest model are the following:

- **Market risk premium**: the difference between market returns and the risk-free rate. According to CAPM, this is the sole driver of stock performance and investors are compensated with higher excess returns only if they hold stocks that present a higher correlation with the market index.

- **Size premium**: the difference between small-cap and large-cap stock returns. This factor is added to capture the tendency of small-cap stocks to outperform large-cap ones.

- **Value premium**: the difference between high book-to-market and low book-to-market stocks. This factor captures instead the tendency of value stocks to outperform growth ones.
- **Profitability premium**; the difference between the returns on diversified portfolios of stocks with robust and weak profitability. This factor captures the tendency of firms with higher profitability to outperform those with lower profitability
- **Investment premium**: the difference between the returns on diversified portfolios of stocks of low and high-investment firms, which we call conservative and aggressive. This factor captures the tendency of firms with more conservative investment strategies to outperform those with more aggressive investment strategies.

The CAPM considers only the first of these factors while the three-factors model the first three.

The mathematical representation of the Fama-French Five-Factor model is the following:

$$\begin{aligned} R_{it} - R_{Ft} = \alpha_i + \beta_{1i}(R_{Mt} - R_{Ft}) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \\ + \beta_{4i}RMW_t + \beta_{5i}CMA_t + \epsilon_{it}, \end{aligned} \tag{3.2}$$

where R_{it} is the return of stock i at time t, R_{Ft} the risk free rate at time t, R_{Mt} the return of the market portfolio, α_i the abnormal return of stock i, β_{1i}, β_{2i}, β_{3i}, β_{4i}, β_{5i}, its factor loadings and ϵ_{it} the error term. This model, which is considered more robust than the CAPM, as it captures multiple sources of systematic risk, has been chosen as the main one for the estimation of normal returns in this study. While the three-factor model and the CAPM are implemented to confirm the strength of the findings.

There are several characteristics of daily excess returns that may have an impact on the validity of the results obtained through the three models, which are simple linear regressions estimated by the OLS method. Firstly, daily security returns are not normally distributed. Their distribution is usually more fat-tailed compared to a normal one, and the same holds true for daily excess returns. This, however, does not affect the estimates of the returns, because the OLS estimator remains the best linear unbiased one also in this case. Non-normality of returns may still cause problems when testing the results, but for test statistics, convergence to the normal asymptotic distribution is quick (MacKinlay, 1997). Moreover, Brown and Warner (1985), in their 1985 study where they analyze the statistical properties of event studies carried with daily observations instead of the previously used monthly ones, state that " departures from normality are less pronounced for cross-sectional mean excess returns than for individual security excess returns, as would be expected under the Central Limit Theorem." and that this applies already for sample sizes of 50 securities. Much smaller than the ones analyzed in this study. In conclusion, the non-normality of returns does not have a remarkable impact on the event study methodology. Secondly, there may be cross-sectional dependence between the errors if the factors in the various models do not capture all returns' comovements. When there is a positive non-zero correlation between errors, not accounting for it, results in a systematic underestimation of the variance of the mean excess return, which leads to rejecting the null hypothesis too many times. But again, Brown and Warner (1985) state that accounting for dependence in the cross-section, really brings better results only in special cases, when it is reasonable to assume that the dependence is strong. Otherwise, for a small degree of dependence, as it is in the case of this study, where event dates are not clustered, ignoring the dependence leads to little bias in the estimation of the variance, while accounting for it, could actually cause harm to the statistical power of the tests. Finally, there may be autocorrelation among the residuals of the regressions. But different studies assert that it is not a problem for the event study methodology, and that the extra work done to account for it, for example by using HAC standard errors, does not seem to strengthen significantly the results (Brown and Warner, 1985) (Henderson, 1990).

Once expected returns have been estimated and actual ones collected, abnormal returns can be measured and analyzed. The abnormal return $AR_{i,t}$ represents the difference between the actual return experienced by stock i in day t of the event window and its expected return under normal circumstances.

$$AR_{i,t} = R_{i,t} - \mathbb{E}[R_{i,t}], \tag{3.3}$$

where $R_{i,t}$ represents the actual return and $\mathbb{E}[R_{i,t}]$ the expected return calculated using the chosen estimation model.

the sample variance of the abnormal returns of stock i, is estimated as:

$$\hat{S}_i^2 = \frac{1}{m-k} \sum_{t=T_0}^{T_1} AR_{it}^2, \tag{3.4}$$

where m is the number of observations of stock prices of i for the estimation window; k is the number of variables used to estimate the normal returns ,including the intercept (so k=6 for the Fama-French 5 factors model); T_0 is the time in which the estimation window begins and T_1 the time in which it ends. The sample variance of abnormal returns should also include a second component, common to all observations, to account for the sampling error in the alphas and the betas, which would cause autocorrelation between abnormal returns. This component though goes to zero for a large enough estimation window, as they are in this study. $\hat{S}_i^2$ is though a good estimator, as the abnormal returns are independent through time. Under the null hypothesis H_0, for which the event has no impact on the behavior of stock returns, abnormal returns are thus normally distributed $AR_{it} \sim \mathcal{N}(0, \hat{S}_i^2)$.

$(T_2 - T_1)$ abnormal returns were computed for n stocks, with T_1 being the event date and T_2 the day on which the event window ends. There are two ways to aggregate them, through time and across securities. The cumulative abnormal return $CAR(T_1, T_2)$ is the accumulation through time of the abnormal returns of a single stock i during the event window:

$$CAR_i = \sum_{t=T_1}^{T_2} AR_{it}, \tag{3.5}$$

and, asymptotically, as the length of the event window L increases, has variance equal to:

$$\sigma_i^2(T_1, T_2) = L\hat{S}_i^2, \tag{3.6}$$

where L is the length of the event window. So, the distribution of the CARs under H_0 is $CAR_i(T_1, T_2) \sim \mathcal{N}(0, \sigma_i^2(T_1, T_2))$.

Returns could have been aggregated across securities instead of through time, and the average abnormal return for each day of the event window could have been calculated. However, given that the goal of the analysis is to accumulate abnormal returns both through time and across securities to assess the overall average impact on returns of ESG rating change announcements, the order in which the two aggregations occur is not relevant. Computing cumulative abnormal returns for each stock and then averaging them across the sample or computing average abnormal returns for each day of the event window and then averaging them for its length, leads to the same result.

Making this aggregation across two dimensions, the important assumption that there is no clustering, hence that there is no overlap between the event windows of the included securities, which is realistic in this case, has to be made in order to prevent nice properties of the returns that make testing easier. In particular, this assumption implies that abnormal returns and cumulative abnormal returns are independent across securities.

The result of this two-dimensional accumulation of returns is the cumulative averaged abnormal return (CAAR), which is nothing else than the average of the CARs:

$$CAAR = \frac{1}{N}\sum_{i=1}^{N} CAR_i, \tag{3.7}$$

and its variance, needed to conduct a test on the significance of this result, is:

$$\sigma_{CAAR}^2 = \frac{\sum_{i=1}^{N} \sigma_i^2(T_1, T_2)}{N^2}, \tag{3.8}$$

which is the sum of the variances of the CARs, based on the assumption that cumulative abnormal returns are independent across securities. In the end $CAAR \sim \mathcal{N}(0, \sigma_{CAAR}^2)$.

Figure 3.6 is a graphical representation of what is the goal of the analysis. From the day of the event (0 on the x-axis), CARs clearly start to increase for the firm that has experienced a positive change in the rating and to decrease for the one that experienced a negative one instead, while there is no particular activity regarding the stock price of the firm whose rating remained unchanged. Of course, the three cases chosen for the representation have been cherry-picked from the sample, and are just an illustration of how one could expect prices to move after a significant event. In reality, it is rare to find such a clear response to new information, nor it is easy to understand what part of the change can be attributed to it.

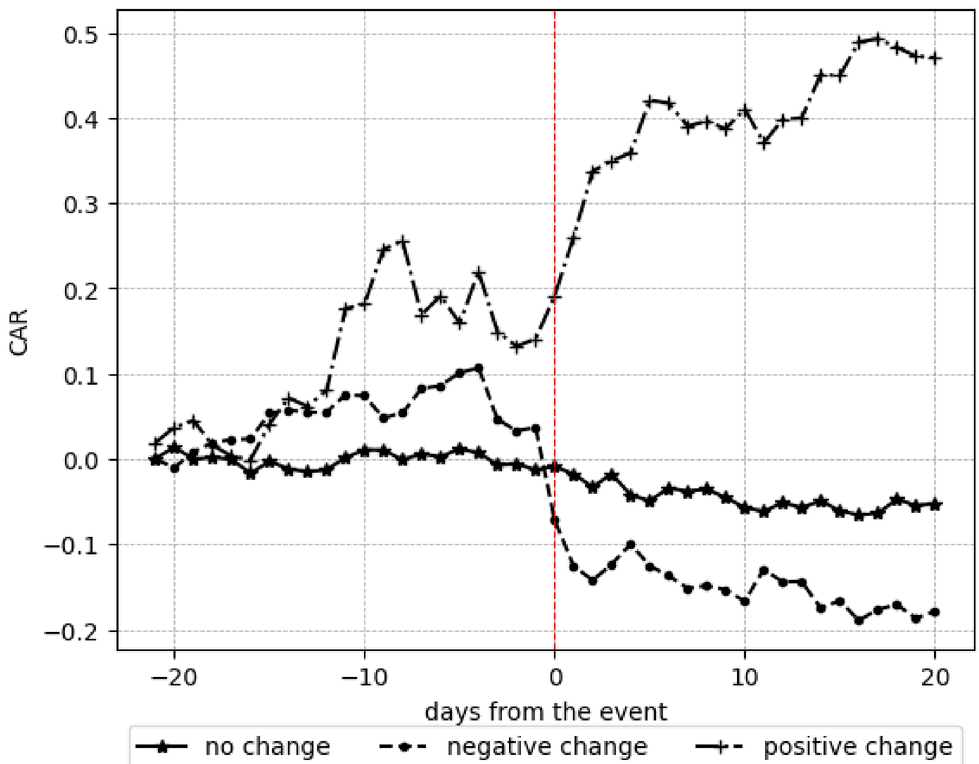

Figure 3.1: Plot of cumulative abnormal returns

This figure represents the CARs for three stocks, one from the sample that experienced a positive rating change, one from the other that experienced a negative one and one from the sample where ratings were confirmed. This is an example of what one would expect to observe in an event study: there is a clear abnormal behavior that follows the event date.

3.2 Tests

The results obtained using the methodology described in the previous section need to be tested. It is reasonable to question if the obtained CAAR is significantly different from zero, and set the null hypothesis accordingly:

$$\mathbb{H}_0 : \mathbb{E}[CAAR] = 0,$$

a t-test can be conducted to decide to reject or less the null hypothesis. The t-value is:

$$t_{CAAR} = \frac{CAAR}{\sigma_{CAAR}}. \tag{3.9}$$

The approximate null distribution of the t-value is a t-distribution with N-1 degrees of freedom t $\sim t_{N-1}$. the t-value is a random variable because abnormal returns are measured with error, which stems from the imprecision in the prediction of the unconditional expected returns and from the fact that individual firms' realized returns during the event window are also affected by reasons unrelated to the event, and this component of the abnormal returns does not average necessarily to zero in the cross-section.

3.3 Definition of event and estimation windows

After having defined the event of interest, the announcement of the rating increase or decrease, the next step to measure the effect of it on stock prices, is to specify the period over which the returns of the securities involved in it will be examined, the event window.

Usually, the event window is not limited to the day of the announcement and expands to multiple days, including at least the day after it, to capture price effects that occur after the stock market closes on the event date (MacKinlay, 1997). The goal is to include in the event window all the days on which price movements that are related to the event take place. Thus, the event window can include days prior to the event date if there is suspicion of information leakage or days far away from it if the market is expected to adjust over a longer period of time. The trade-off, when

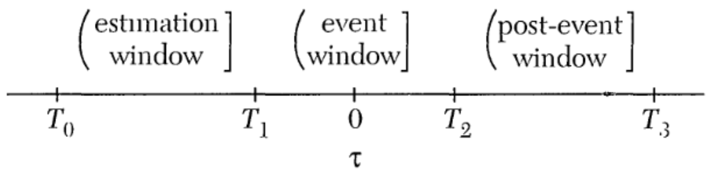

Figure 3.2: Time line for an event study. ***Source: Mackinlay, 1997***
This is the timeline of an event study. $\tau = 0$ is the event date. The estimation window goes from an arbitrary day T_0 to T_1, usually the day before the event. The event window goes from T_1 to T_2 and often comprehends the day of the event and some of the days after it. In some cases, as in the image, also days before the event to account for possible information leakages. The post-event window, which is not used in this study, and rarely is, goes from T_2 to T_3.

choosing the length of the window, is between increased precision, obtained using a longer event window that includes more of the changes caused by the event, and the inclusion of price changes caused by unrelated events, that can compromise the computation of abnormal returns. If the market is assumed to be semi-efficient, meaning that all public information is readily incorporated into prices (Fama, 1970), then the event window can be pretty small, since the market immediately prices the new information.

In this study, different event windows have been utilized. The shorter one is the one that comprehends the day of the event and the 5 trading days following it. It is represented by the notation [0,5]. Also, two longer windows were utilized, the [0,11] and the [0,21] ones, to account for the ”softness” of the information, which may make it more difficult to analyze the impact of changes, and consequently need more time to be utilized for trading. These lengths are also useful to capture the reaction of sustainable investors, who may adjust their portfolios due to investment strategies such as screening rules. Action that may possibly take longer than 5 days to implement. Also, an event window that includes five days before the event, the [-5,5] one, was utilized to account for possible leakages of information before the announcement.

In order to compute the abnormal return, that is the difference between the actual ex-post return of a stock over the event window and its normal return, which is the expected return conditioned on the event not taking place, one needs to define an estimation window over which collect data. The estimation window usually includes several days prior to the event and excludes the day of the event to prevent it from influencing the normal parameter estimates (MacKinlay, 1997). Estimation periods range from 100 to 300 days and face the trade-off between the precision of the estimates of the parameters and the risk of them being ’out of date’ (Armitage, 1995). Corrado and Zivney (1992) compare results from three test statistics using pre-event estimation periods of 239, 89 and 39 days and they found out that they were virtually unaffected by an estimation period of 89 instead of 239 days, and that a 39-day period produced ‘only a slight deterioration of performance’, so 100 days or more seems a safe choice for the estimation window. In this study, a length of 110 days is given to the window, with the last day being the one before the event.

Chapter 4

Empirical Analysis

4.1 Data

ESG data from the Intangible Value Assets database of Morgan Stanley Capital International (MSCI), which is considered the largest ESG data vendor by the investment community (Welch and Yoon, 2020), is a widely used source of information in academic research. In this study, aggregate ESG ratings, as well as ratings regarding each of the three E, S, and G pillars, are taken from this database, whose choice was driven by the great influence played by the providing corporation, which is probably more capable than others of influencing markets, and by the availability of the information.

An ESG rating aims to measure a company's ability to manage financially relevant ESG risks and opportunities. MSCI ratings identify industry leaders and laggards according to their exposure to ESG risks and how well they manage those risks relative to peers. Aggregate ESG ratings range from AAA to CCC (MSCI, 2019) and are derived from the Industry-Adjusted Company Score, which is a weighted average of the numerical scores that MSCI gives to each company based on its performance relative to 33 different Key Issues, then translated to a letter according to a predetermined table. These issues regard ten different themes across the three pillars: climate change, natural capital, pollution and waste and environmental opportunities for the Environmental pillar; human capital, product liability, stakeholder opposition and social opportunities for the Social pillar and corporate governance and corporate behavior for the Governance one (MSCI, 2024). These ratings are therefore a quasi-quantitative summary of raw, unstructured, qualitative, and voluminous data of a firm, and can be really valuable for investors. It is also possible though, that, as noted by Glück et al. (2021) ”the aggregation of pillar scores to an overall ESG rating could have a dilutive effect on information quality”. To control for this possibility, changes in the pillar ratings were also analyzed singularly in the aggregate analysis. MSCI pillar ratings are weighted averages of the Key Issues scores underlying each pillar and expressed as decimal numbers ranging from 0 to 10.

An ESG rating event is defined as the exact date on which a full analyst review of a firm’s performance relative to ESG factors and the summarizing rating is published. This usually happens once a year, while updates regarding single pillars’ ratings are published every month.

In the analysis of the aggregate ESG variable, upgrades are positive changes of at least one letter in the aggregate ESG rating, while decreases are the opposite. For single pillar ratings, an upgrade is instead an increase in the numerical rating of at least 1.0, a decrease the opposite, while changes that in absolute value are smaller than 1.0 are considered confirmations of the previous analysis.

The scrutiny of the combined ESG scores for every company in Europe for which data were available highlighted 2,850 rating increases and 1,085 rating decreases, affecting a total of 1,236 firms. Announcements where the aggregated ESG score did not change are of course much more numerous, as illustrated in Figure 4.1. This discrepancy in numbers between changes of opposite signs, which persists also in the sector-specific analysis, may be due to increased attention of companies to this aspect of their operations or to a difficulty among agencies in releasing negative information about firms’ performances in this field.

For the Environmental pillar, there were 2,177 upgrades and 1,309 downgrades, for the Social pillar 1,759 upgrades and 1,670 downgrades, and for the Governance pillar 2,311 upgrades and 2,592 downgrades. For single pillar ratings, that are updated monthly also if they do not change very frequently, the proportion of times in which the rating remained the same is even higher than the 80.41% of the aggregate rating.

The choice of considering every company in Europe for which relevant data were available, and not only the most significant ones, like those comprehended in the *Eurostoxx600* index, was driven by the willingness to have the largest possible sample to make inferences from.

At first, for both the combined ESG scores and each of the single pillars' ones, companies were divided into three datasets, one containing the ones that suffered an ESG rating decrease, one those that experienced an increase and the other the ones that saw their rating confirmed. For each one of these companies, the stock returns needed to carry on the analysis were retrieved from Yahoo Finance. The same analysis described above, limited to the combined ESG variable, also runs distinctly on some of the most relevant sectors: energy, banking, and utilities. More precisely, these macro sectors have been constructed by aggregating similar smaller ones, given that MSCI groups companies also in particularly small niches, that contain less than 50 announcements each. The energy macro sector contains the MSCI IVA industries ''Energy Equipment & Services', 'Oil & Gas Exploration & Production', 'Integrated Oil & Gas','Oil & Gas Refining, Marketing, Transportation & Storage', and highlighted 74 upgrades, 32 downgrades and 397 confirmations of the previous rating. The banking macro sector comprehends the industries 'Banks', 'Asset Management and Custody Banks', 'Investment Banking and Brokerage' and revealed 428 upgrades, 122 downgrades and 1346 confirmations, while the statistics for the utility macro sector, which included only the industries 'Utilities' and 'Wireless communication services', are the following: 122 upgrades, 555 confirmations and 46 downgrades. In all the sub-samples the positive news are more numerous than the negative ones, and both are significantly less than the announcements of confirmations of the rating, confirming the trend noticed in the aggregate database.

The analysis of the returns was actually made on samples with sizes smaller than the ones declared above, but big enough to make inferences. This was sometimes due to the lack of information in the MSCI dataset, which had missing ISINs and tickers, whose absence made it impossible to find the correct prices for certain companies, and sometimes to the impossibility of finding a link between the two datasets for certain stocks.

4.2 Summary statistics

The original dataset on which the analysis was performed contained 20,091 IVA rating announcements from MSCI on European companies (including GB), made between the dates 2016-05-27 and 2023-08-29. 16,156 of these (80.41%) are actually confirmations of the previous rating, while 2,850 (14.18%) are announcements of an increase and 1,085 (5.4%) of a decrease. The distribution of rating changes is illustrated in Figure 4.1.

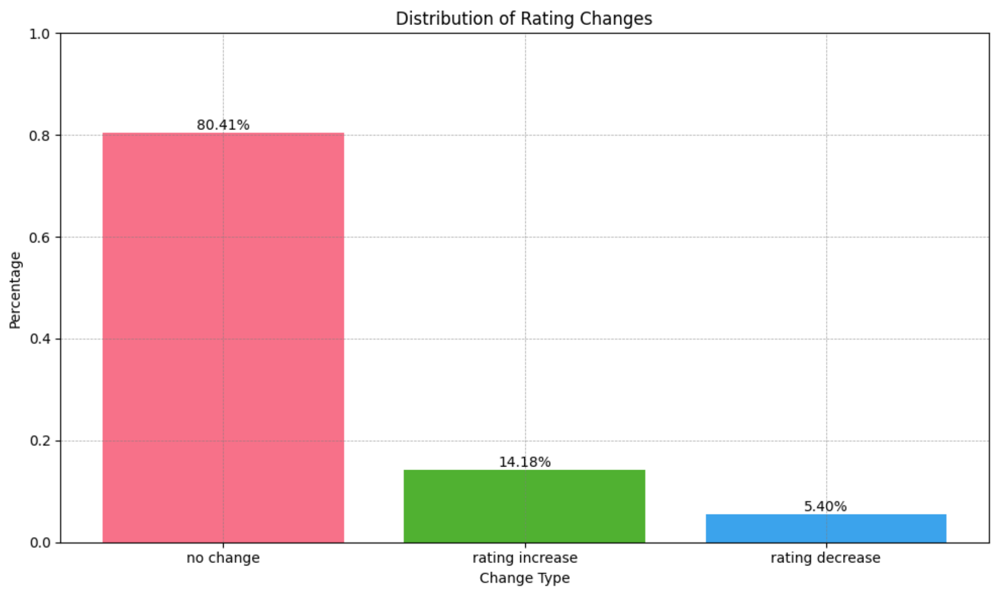

Figure 4.1: Distribution of Rating Changes
This figure shows how the confirmations of ratings are distributed by sectors. These percentages are similar to those in the overall MSCI sample, so there is nothing in particular to notice.

MSCI grouped companies into 70 different sectors, denominated 'IVA INDUSTRIES'. As one can see in figures 4.2, 4.3 and 4.4, which show the distributions of rating changes per industry in the various scenarios, "Banks" is by far and in any case the most represented one with more than 10% of the firms under scrutiny falling into this category. "Real Estate Management & Services" is always second, followed by "Utilities", "Software and Services" and others depending on the direction of the announcement. These percentages are similar to the proportion of each industry in the whole sample, so there is nothing impressive to notice. An interesting fact emerging from this analysis is that "Media & Entertainment" shows up as one of the industries with more downgrades (4th with 4.1% of the announcements in this category) while not being even present among the 15 most represented industries in the upgrade case, meaning that probably firms in this industry saw their rating go down much more than up during the period considered.

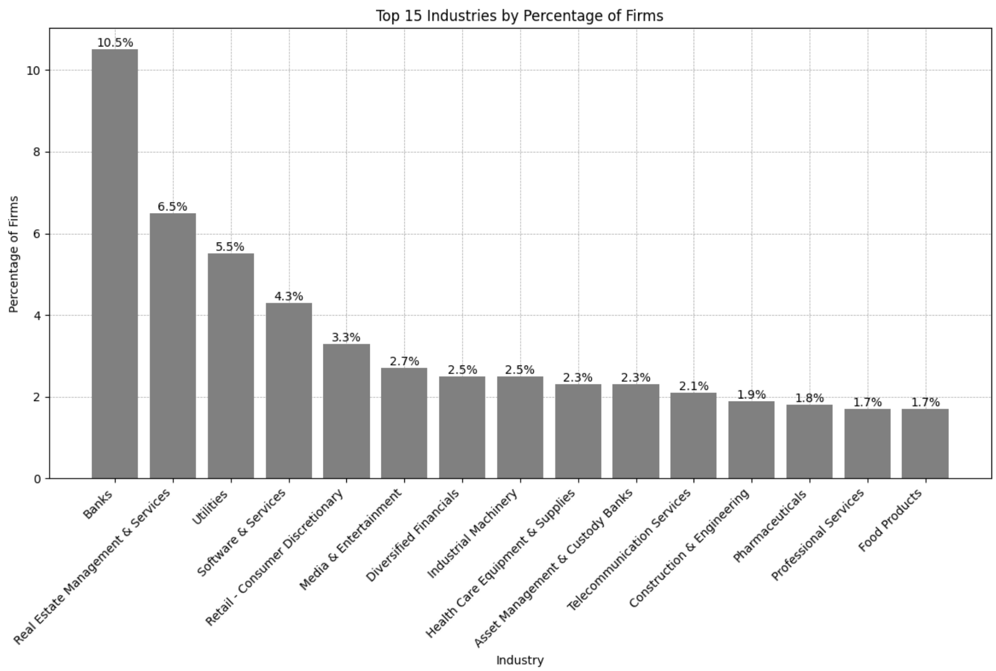

Figure 4.2: Distribution of rating changes in case of confirmation of the rating
This figure shows how the confirmations of ratings are distributed by sectors. The banking sector is clearly the most represented, followed by Real Estate Management and Services, Utilities and Software and Services. These percentages are in line with those of the overall MSCI sample.

Rating changes affected companies from all of the 27 states of the European Union and from Great Britain, which is actually where the highest percentage of firms in the sample, 29.86%, is domiciliated, followed by France (10.11%), Germany (8.74%), Sweden (8.62%), Netherlands (8.22%), Luxembourg (6.11%), Italy (5.39%.) and the others, as illustrated in figure 4.5.

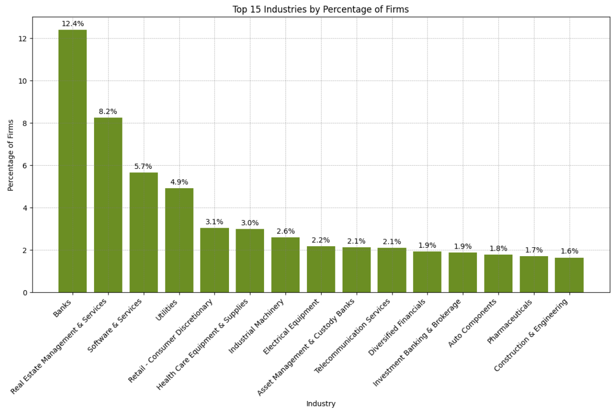

Figure 4.3: Distribution of rating announcements per industry in case of positive change
This figure shows how positive changes in the ratings are distributed by sectors. The distribution is very similar to the one of the confirmations, suggesting that no industry saw significantly more upgrade reviews than confirmations.

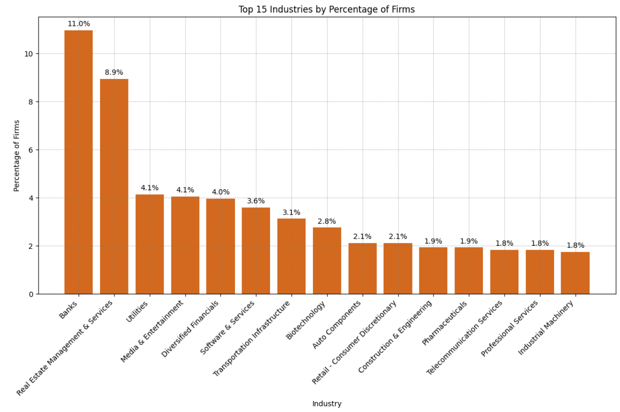

Figure 4.4: Distribution of rating announcements per industry in case of negative change
This figure shows how negative changes in the ratings are distributed by sectors. Stocks in the Media & Entertainment sector represent 4.1% of the total number of stocks in this sub-sample, while this sector is not even among the first 15 most represented in case of positive change or confirmation of the rating. Probably this sector experienced much more negative rating reviews than positive.

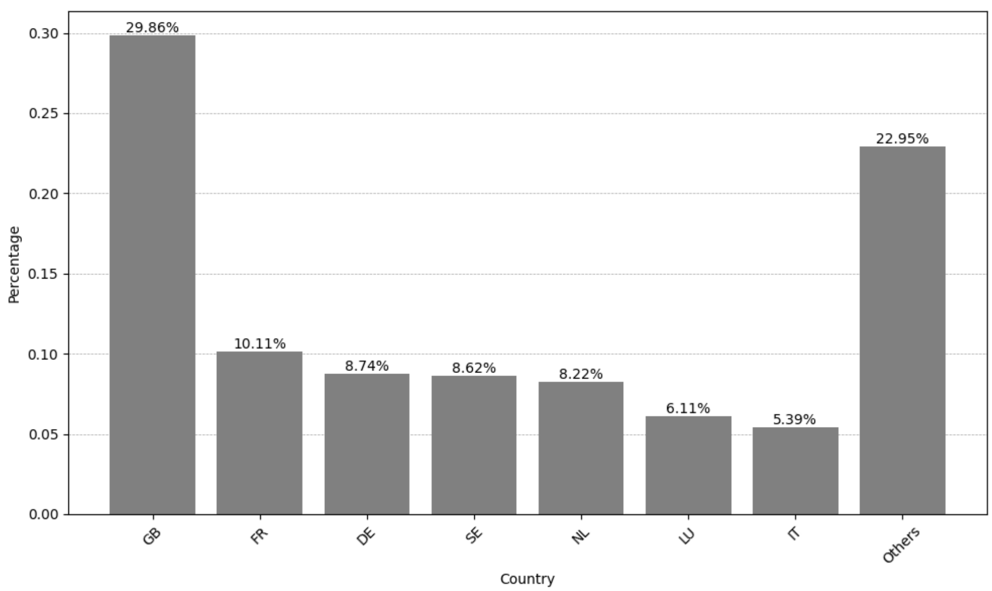

Figure 4.5: Distribution of rating changes per country
The figure shows how combined ESG rating changes are distributed across countries. Great Britain is by far the nation that is domicile to more companies affected by changes, followed by France, Germany and, surprisingly, Sweden, before the others.

4.3 Results

This section presents the findings of the event study on the effects of ESG rating change announcements on stock prices. The analysis examines the Cumulative Average Abnormal Returns (CAARs) for different event windows, considering both the combined ESG score and the scores on the individual pillars. For robustness, three different models are employed to compute normal returns: a one-factor (CAPM), three-factor, and five-factor Fama-French model. After a first analysis of the aggregate data, subsequent were conducted on relevant sub-samples to test the consistency of the patterns observed in the market reaction to the variation in the combined ESG rating and eventually notice whether interesting variations of these were present among them.

The results of the first analysis are presented in Table 4.1. By examining the first event window, which is the shortest one and includes the day of the announcement and the five subsequent trading days, it can be noticed that for aggregate ESG rating changes, positive CAARs are observed across all scenarios (upgrades, downgrades, and confirmations) and all models. However, none of these results are statistically significant at conventional levels. The magnitude of CAARs ranges from 0.52% to 0.97%, with downgrades showing slightly higher returns than upgrades and confirmations.

These findings are somewhat unexpected. The CAARs reflect how the market collectively perceives the value of the announcement of a change in ESG rating and the literature claims the actual and growing importance of these variables, which, as exposed in Chapter 2, give valuable insights about a firm's long-term performance. However, the analysis only reveals mild positive reactions in the short term, not only for upgrades but, more surprisingly, also for downgrades and confirmations, which are nothing more than observable trends without any statistical significance, not even at the 10% level. One possible explanation for this lack of reaction is that European investors are actually still skeptical about the financial materiality of news regarding the combined ESG variable, at least in the short term. Another, less likely interpretation of the results, is that the market had already priced the information, and for this reason, the announcement did not provoke any significant reaction. This would mean that all the information on which the rating is

Table 4.1 shows cumulative average abnormal returns related to increases (+), decreases (-), and confirmations (0) of the ESG combined rating and of the single E, S and G ones. For robustness, for the same event normal returns were estimated using three different models, containing respectively one, three and five of the factors illustrated by E.Fama and K.French in their 2013 paper Fama and French (2014). The applied test statistics are approximately normally distributed. z-values are shown in parentheses. ***, ** and * indicate statistical significance at 1%, 5% and 10% level. The notation $[t_1, t_2]$ means that the results are obtained using an event window that goes from t_1 trading days before the event to t_2 trading days after it.

Table 4.1: Combined CAARs Data

	CAAR [0,5]				CAAR [0,11]				CAAR [0,21]				CAAR [-5,5]			
Model	ESG	E	S	G	ESG	E	S	G	ESG	E	S	G	ESG	E	S	G
1-factor (+)	0.88%	0.92%	0.50%	0.66%	1.47%	1.37%	0.81%	1.33%	2.01%	1.45%	1.25%	2.33%	0.32%	0.53%	0.20%	0.92%
	(0.50)	(0.67)	(0.38)	(0.45)	(0.56)	(0.67)	(0.42)	(0.61)	(0.57)	(0.53)	(0.48)	(0.79)	(0.13)	(0.27)	(0.11)	(0.44)
1-factor (0)	0.53%	1.85%	0.68%	0.60%	0.43%	1.85%	0.84%	0.73%	0.87%	1.83%	1.25%	1.34%	0.42%	0.75%	0.67%	0.56%
	(0.95)	(0.56)	(0.49)	(0.42)	(0.14)	(0.54)	(0.40)	(0.40)	(0.21)	(0.55)	(0.51)	(0.54)	(0.15)	(0.32)	(0.30)	(0.32)
1-factor (-)	0.97%	1.81%*	1.11%	0.69%	1.69%	1.57%	1.76%	1.94%	2.11%	2.51%	2.50%	3.28%	0.54%	1.20%	0.66%	-2.10%
	(0.76)	(1.87)	(0.90)	(0.44)	(0.89)	(1.09)	(0.96)	(0.84)	(0.82)	(1.30)	(1.01)	(1.05)	(0.30)	(0.88)	(0.38)	(-1.06)
3-factor (+)	0.83%	0.87%	0.42%	0.57%	1.37%	1.25%	0.74%	1.25%	1.93%	1.36%	1.15%	2.26%	0.38%	0.53%	0.14%	0.79%
	(0.47)	(0.63)	(0.33)	(0.39)	(0.52)	(0.61)	(0.39)	(0.58)	(0.55)	(0.49)	(0.44)	(0.77)	(0.15)	(0.27)	(0.08)	(0.38)
3-factor (0)	0.52%	1.77%	0.66%	0.51%	0.41%	1.80%	0.79%	0.65%	0.78%	1.61%	1.17%	1.16%	0.46%	0.63%	0.63%	0.48%
	(0.25)	(1.05)	(0.54)	(0.42)	(0.13)	(0.60)	(0.44)	(0.40)	(0.19)	(0.60)	(0.47)	(0.47)	(0.16)	(0.39)	(0.37)	(0.27)
3-factor (-)	0.95%	1.78%*	1.02%	0.38%	1.59%	1.54%	1.59%	1.49%	1.97%	2.43%	2.30%	2.63%	0.56%	1.18%	0.54%	-2.40%
	(0.74)	(1.84)	(0.83)	(0.32)	(0.84)	(1.07)	(0.87)	(0.32)	(0.77)	(1.26)	(0.94)	(0.85)	(0.31)	(0.87)	(0.31)	(-1.06)
5-factor (+)	0.89%	0.86%	0.42%	0.62%	1.42%	1.24%	0.77%	1.28%	1.97%	1.40%	1.15%	2.15%	0.45%	0.53%	0.16%	0.87%
	(0.51)	(0.57)	(0.32)	(0.43)	(0.54)	(0.61)	(0.41)	(0.53)	(0.56)	(0.63)	(0.45)	(0.70)	(0.16)	(0.08)	(0.08)	(0.38)
5-factor (0)	0.57%	1.49%	0.68%	0.49%	0.43%	1.49%	0.82%	0.57%	0.78%	1.30%	1.18%	1.13%	0.49%	0.48%	0.65%	0.59%
	(0.28)	(1.08)	(0.54)	(0.40)	(0.14)	(0.60)	(0.45)	(0.37)	(0.18)	(0.63)	(0.48)	(0.43)	(0.17)	(0.39)	(0.37)	(0.32)
5-factor (-)	0.91%	1.69%*	1.02%	0.48%	1.46%	1.44%	1.61%	1.41%	1.93%	2.31%	2.25%	2.56%	0.50%	1.14%	0.51%	-2.30%
	(0.72)	(1.75)	(0.83)	(0.32)	(0.77)	(1.01)	(0.88)	(0.62)	(0.75)	(1.20)	(0.92)	(0.83)	(0.28)	(0.83)	(0.31)	(-1.06)

constructed was actually already public and that the work of the providing agencies (MSCI in this case) does not provide any added value to society. As illustrated in Chapter 1 though, the ESG rating is much more than the mere summation of ESG news and the process that produces it takes also non-public information as inputs, hence it is probably incorrect to dismiss its value based on this analysis alone.

The reason for a deeper scrutiny of what composes the combined ESG rating was the chance of a dilutive effect on information quality caused by the aggregation of the E,S, and G ratings in a single one. Indeed, when examining individual pillars using the [0,5] event window, the Environmental (E) pillar stands out. Downgrades in this case show the largest and only statistically significant CAARs, which range from 1.69% to 1.81% and are significant at the 10% level across all models. Again, this is somewhat unexpected and counter-intuitive, but, as noted in Chapter 2, the effect that good ESG performance has in the short term is ambiguous, given that can actually decrease short-term financial performance or be driven by egoistic purposes of managers. The Social (S) and Governance (G) pillars also show positive CAARs, but these are not statistically significant. Notably, for the Social pillar, downgrades show slightly higher returns compared to upgrades and confirmations.

Extending the analysis to a longer [0,11] event window, it is clear the persistence of the patterns observed above. CAARs remain positive but statistically insignificant across all scenarios for aggregate ESG rating changes. For individual pillars, CAARs related to changes in the Environmental one remain positive but lose statistical significance compared to the [0,5] window, signaling the efficiency of the market in promptly pricing the information. The Social and Governance pillars show increased CAARs compared to the shorter window, particularly for downgrades, but these results remain statistically insignificant. These results are interesting, considering that the introduction of a longer window was driven by concerns regarding the softness of the information, and consequently the need for a longer period of time needed to adjust to it.

The longest [0,21] event window, confirms what is seen in the [0,11] one. It shows still an increase in the magnitude of all of the CAARs. For aggregate ESG rating changes, CAARs range from 0.78% to 2.11%, with downgrades consistently showing the highest returns, followed by upgrades and, as expected, confirmations. However, these results remain statistically insignificant. Among individual pillars, the Governance pillar shows the largest increase in CAARs, particularly for downgrades, with positive CAARs of 3.28%, 2.63%, and 2.56% computed with the three different models that are the highest in this analysis. Yet, none of these reaches statistical significance at any conventional level.

Lastly, in order to examine potential information leakage, which is surely more likely in the developing ESG world than in the heavily regulated one of corporate finance, a symmetric [-5,5] event window, which encompasses five trading days before and after the announcement, was introduced. For aggregate ESG rating changes, CAARs are positive but smaller in magnitude compared to any of the other windows and remain statistically insignificant. The same is true for the Environmental and Social pillars. Interestingly, for the Governance pillar, downgrades show negative CAARs ranging from -2.10% to -2.40%. They are the only negative ones in all of the analyses, although they are not statistically significant. Upgrades and confirmations for this pillar remain positive.

These results provide insights into the market's reaction to ESG rating change announcements. Across all event windows and rating change types, predominantly positive CAARs are observed, suggesting a general positive market reaction to ESG rating announcements. Surprisingly, downgrades consistently show higher CAARs compared to upgrades and confirmations, particularly for individual pillar ratings. The only statistically significant result in the aggregate data is a positive market reaction that follows a downgrade in the Environmental pillar, found only for the [0,5] event window. This result is consistent across the three different estimation models. There is also evidence of a potential delayed market reaction to ESG rating changes, as CAARs generally increase in magnitude as the event window expands. However, the lack of statistical significance in most cases indicates that while there are observable trends, the market reaction to ESG rating changes is not consistently strong enough to rule out random variation.

Moreover, the analysis of the [-5,5] window does not provide strong evidence of information leakage before the announcements. It is important to notice that results are generally consistent across the three models used to compute normal returns. This strengthens the robustness of the findings.

The only case in which the market reacts more markedly to one of these announcements is in correspondence with a decrease in the rating in the Environmental pillar. In this case, actual returns are consistently higher than predicted ones in the first five trading days following the event. The fact that this significance fades for larger event windows is actually evidence of the efficient market discussed above since it takes little time to incorporate this new relevant information into prices.

Overall, these findings, which mostly lack statistical significance, suggest that the market does not react markedly to changes in ESG variables. This is likely not indicative of market inefficiency, but rather of the fact that ESG information may not be as relevant or material to investors as other types of financial information, at least in the short term. According to the efficient market hypothesis, if ESG information were truly value-relevant, it would be more promptly and significantly reflected in stock prices, and this would have been reflected in CAARs statistically significant from 0 in this study.

The results of the second analysis, performed over the sub-samples of companies in the energy, banking, and utilities sectors, highlight more interesting findings, presented respectively in tables 4.2, 4.3, and 4.4.

Stocks of companies in the energy sector display a behavior that is in contrast to the evidence provided by the aggregate sample. The upgrade of the combined ESG score did not seem to produce any relevant reaction in the market, apart from a non-statistically significant positive trend. Now instead, it generates a negative reaction of roughly 100 basis points consistent across all the event windows, which is statistically significant at the 5% level according to all three models. Also, the [-5,5] event window, which never produced interesting results on the aggregate data, signals the same evidence, suggesting possible information leakage regarding this kind of information in this sector. There are a few exceptions to these findings, for instance, the 5-factor model still shows a negative CAAR of -0.76% for the [0,11] event window, but it is not statistically significant, while the negative CAARs observed in the [0,21] event window only reach the 10% threshold.

Table 4.2: CAARs of stocks in the energy sector following a change in the combined ESG rating

Table 4.2 shows the cumulative average abnormal returns (CAARs) for different event windows for the sub-sample of companies operating in the Energy sector. Results are shown for upgrades (+), downgrades (-), and confirmations (0) of the combined ESG rating across models with one, three, and five factors. The z-values are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The notation $[t_1, t_2]$ means that the results are obtained using an event window that goes from t_1 trading days before the event to t_2 trading days after it.

	[0,5]	[0,11]	[0,21]	[-5,5]
1-factor model				
(+)	-0.90%**	-1.14%**	-1.23%*	-1.07%**
	(-2.57)	(-2.18)	(-1.75)	(-2.16)
(0)	1.51%	2.01%	4.76%	1.33%
	(1.01)	(0.91)	(1.59)	(0.91)
(-)	2.66%***	3.02%***	5.12%***	1.11%***
	(10.97)	(8.40)	(10.55)	(3.33)
3-factor model				
(+)	-0.86%**	-1.12%**	-1.30%*	-1.04%**
	(-2.45)	(-2.13)	(-1.83)	(-2.07)
(0)	1.54%	2.04%	4.78%	1.38%
	(1.04)	(0.93)	(1.60)	(0.66)
(-)	2.51%***	3.02%***	4.75%***	0.49%
	(10.54)	(8.46)	(9.88)	(1.45)
5-factor model				
(+)	-1.05%***	-0.76%	-1.61%**	-1.40%**
	(-2.96)	(-1.46)	(-2.28)	(-2.79)
(0)	1.37%	1.80%	4.59%	1.26%
	(0.92)	(0.82)	(1.54)	(0.60)
(-)	2.27%***	2.14%***	3.31%***	0.80%**
	(9.48)	(6.03)	(6.91)	(2.35)

When a downgrade in the combined ESG rating is announced instead, a positive reaction, as already seen in the aggregate data, follows, with the differences that, for this sub-sample, it is generally larger (ranging from 2.14% to 5.12% not considering the [-5,5] event window) and statistically significant, at the 1% level and beyond, across all models and event windows. The exception here is the [-5,5] event window, where less negative CAARs, significant at the 5% level, can be observed, which may suggest a mild unexpected market reaction starting already in the days leading to the announcement. As expected, CAARs following confirmation of the previous rating are never significant nor noteworthy.

Results regarding the energy sector are somehow counterintuitive: increases in a variable that is supposedly a signal for better financial performance, lead to negative short-term returns for the stocks, while downgrades produce the opposite effect. A possible explanation may be that in a sector where companies work for Oil and Gas Exploration, Refining, Transportation, and so on, money spent on 'greener' projects can not yield the highest short or medium-term return. Consequently, an upgrade in the ESG variable may actually be interpreted as a negative signal for shareholders primarily focused on the more immediate profitability that an investment in historically lucrative, but less sustainable ventures, may achieve.

In contrast to the energy sector's clear-cut reaction to changes in the combined ESG rating, the banking sector exhibits a milder response, that still differs though from that of the market taken as a whole.

While the energy sector analysis reveals negative and statistically significant CAARs following an increase in the ESG variable across the majority of event windows and models, the banking sector shows CAARs do not follow any particular trend with none of them ever reaching any common significance level. This suggests that the market may perceive ESG improvements differently across the two sectors. Interestingly, both sectors demonstrate positive CAARs in response to ESG rating downgrades, although the magnitude of them varies. The energy sector shows a clear-cut response, while in the banking one CAARs are still big in absolute terms but relatively smaller, ranging from 1.13% to 1.80%. This pattern is consistent across all event windows and models, with the exception of the [-5,5] window, introduced to signal possible information leakage that is never significant nor exhibits any clear trend. Also in this case, confirmation of the previous rating, never produces any noteworthy reaction.

The utility sector was the last one to be analyzed. It shows similar results to the energy sector for what concerns downgrades, with highly significant positive CAARs in every event window,

Table 4.3: CAARs of stocks in the Banking sector following a change in the combined ESG rating

Table 4.3 shows the cumulative average abnormal returns (CAARs) for different event windows for the sub-sample of companies operating in the Banking sector. Results are shown for upgrades (+), downgrades (-), and confirmations (0) of the combined ESG rating across models with one, three, and five factors. The z-values are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The notation $[t_1, t_2]$ means that the results are obtained using an event window that goes from t_1 trading days before the event to t_2 trading days after it.

	[0,5]	[0,11]	[0,21]	[-5,5]
1-factor model				
(+)	0.19%	0.35%	0.71%	-0.44%
	(0.32)	(0.40)	(0.60)	(-0.52)
(0)	0.34%	0.73%	0.87%	-0.24%
	(0.37)	(0.55)	(0.48)	(-0.19)
(-)	1.16%***	1.30%***	1.70%***	0.28%
	(4.68)	(3.54)	(3.42)	(0.80)
3-factor model				
(+)	0.01%	0.20%	0.52%	-0.47%
	(0.02)	(0.23)	(0.44)	(-0.56)
(0)	0.33%	0.68%	0.79%	-0.14%
	(0.36)	(0.51)	(0.44)	(-0.11)
(-)	1.13%***	1.28%***	1.68%***	0.30%
	(4.56)	(3.49)	(3.38)	(0.87)
5-factor model				
(+)	0.06%	0.27%	0.61%	-0.38%
	(0.11)	(0.31)	(0.51)	(-0.45)
(0)	0.44%	0.83%	0.86%	0.01%
	(0.49)	(0.62)	(0.48)	(0.00)
(-)	1.16%***	1.41%***	1.80%***	0.37%
	(4.65)	(3.82)	(3.61)	(1.04)

[-5,5] included, confirmed by all three models. The magnitude of these returns increases with the length of the event window, with CAARs in the [0,21] window being always the largest, reaching an enormous 5.66% when computed with the 1-factor model, which more than doubles the 2.76% observed in the shorter window.

This suggests that the market immediately reacts to the new information, as one can see from the significant CAARs in the shorter window, but then takes time to fully incorporate it into prices. The explanation for this reaction may be similar to that of the energy sector. Utilities often include large companies for which innovation can be difficult, and that often offer steady dividends to their shareholders. They may see a downgrade in this variable as a signal of the company focusing on a safer source of revenues and hence as positive news.

In contrast with the energy sector though, which showed significant negative CAARs following an upgrade in the ESG variable, the utility sector highlights instead a positive and statistically significant (at the 1% level) market reaction in this scenario. As noted for downgrades, this price movement seems to last and to increase over time, with CAARs for the [0,21] windows being always greater than 3% according to all models and CAARs for the shortest one, [0,5], being around 75 basis points. Interestingly, the [-5,5] window does not produce interesting results, except for a CAAR of 0.60% significant at the 10% level captured only by the three-factor model, while it showed significant results for downgrades, suggesting that negative news may be more anticipated than positive ones. Also in this case, rating confirmations do not produce relevant results.

The utility sector's seemingly paradoxical positive market reaction to both upgrades and downgrades underscores the complexity of the investor sentiment toward the ESG metric. While the literature asserts the positive correlation between financial and ESG performance in the long run, the effects of variations in these kinds of variables in the short term seem to be more complex and are open to different interpretations.

In conclusion, the market as a whole does not show any particular reaction to the variation of the combined ESG score, nor to changes in the Social and Governance ones. On one hand, the aggregate analysis only revealed positive trends following changes in both directions, which are not strong enough to rule out the hypothesis of being the result of random variations. On the other hand, when changes in single pillars were analyzed, to investigate whether this lack of reaction was due to a loss of information caused by the aggregation of the different ratings in a single one, it revealed the different sensitivity of the market to the Environmental variable, whose downgrades seem to cause an unexpected positive market reaction that is promptly incorporated into prices.

When analyzing distinctly three of the most relevant sectors instead, the combined ESG score demonstrated to be, even alone, an important vector of information. Its changes led to a variety

Table 4.4: CAARs of stocks in the utility sector following a change in the combined ESG rating

Table 4.4 shows the cumulative average abnormal returns (CAARs) for different event windows for the sub-sample of companies operating in the Utility sector. Results are shown for upgrades (+), downgrades (-), and confirmations (0) of the combined ESG rating across models with one, three, and five factors. The z-values are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The notation $[t_1, t_2]$ means that the results are obtained using an event window that goes from t_1 trading days before the event to t_2 trading days after it.

	[0,5]	[0,11]	[0,21]	[-5,5]
1-factor model				
(+)	0.71%***	2.35%***	3.24%***	0.34%
	(2.86)	(6.43)	(6.57)	(0.98)
(0)	0.49%	0.57%	0.91%	0.40%
	(0.97)	(0.75)	(0.90)	(0.56)
(-)	2.76%***	3.02%***	5.66%***	2.46%***
	(16.58)	(12.23)	(17.00)	(10.44)
3-factor model				
(+)	0.74%***	2.42%***	3.36%***	0.60%*
	(3.00)	(6.64)	(6.85)	(1.74)
(0)	0.48%	0.59%	0.99%	0.42%
	(0.96)	(0.79)	(0.98)	(0.58)
(-)	2.31%***	2.59%***	4.83%***	2.19%***
	(13.96)	(10.54)	(14.57)	(9.33)
5-factor model				
(+)	0.76%***	2.44%***	3.59%***	0.52%
	(3.10)	(6.70)	(7.32)	(1.51)
(0)	0.54%	0.59%	1.00%	0.43%
	(1.07)	(0.79)	(1.00)	(0.60)
(-)	2.52%***	2.84%***	5.15%***	2.69%***
	(15.18)	(11.54)	(15.53)	(11.47)

of reactions, that differed in size and direction across the energy, banking, and utility sectors. Surprisingly, as may be suggested by the results of the first analysis, across all three sub-samples announcements of downgrades resulted in positive and statistically significant CAARs, also for larger event windows, suggesting that in these sectors, in contrast with what the results showed in Table 4.1, this kind of information takes a while to be priced. A completely new finding was the positive reaction to upgrades in the ESG score in the utility sector, and the negative one to the same kind of information for stocks of energy companies.

Chapter 5

Conclusions

This study examined the market reaction to ESG rating announcements for European companies using an event study methodology. The first analysis, performed on the aggregate data, investigated both combined ESG ratings and individual Environmental, Social, and Governance pillar scores, while the second one, which was sector-specific, only focused on the combined score.

The results of the aggregate analysis show a nuanced market reaction to ESG rating changes. Predominantly positive Cumulative Averaged Abnormal returns were observed following any kind of rating announcement. They were computed across various event windows and confirmed by three different estimation models. However, the majority of these results lacked statistical significance, suggesting, more than an inefficiency, that the market does not react markedly to changes in these variables.

Interestingly, downgrades showed consistently higher CAARs compared to upgrades and confirmations, particularly for individual pillar ratings. The most notable finding was a statistically significant positive market reaction following downgrades in the Environmental pillar, observed in the [0,5] event window, which includes the day of the event and the five trading days that follow it. This was consistent across all three estimation models.

The lack of statistical significance in most cases suggests that while there are observable trends, the market reaction to ESG rating changes is not consistently strong enough to rule out random variation. These findings are probably not indicative of market inefficiency but rather suggest that ESG information may not be as value-relevant to investors as other types of financial information, like credit ratings changes.

Moreover, the only instance of a statistically significant market reaction, provided some evidence of market efficiency, as this significance fades for longer event windows, indicating a rapid incorporation of information into prices.

The sector-specific analysis led to distinct results. Downgrades of the combined ESG rating were consistently followed by positive and statistically significant CAARs, that increased as the event window widened. Upgrades were welcomed differently across sectors, causing negative CAARs in the energy sector, positive ones in the utility sector, and no particular reaction for banks.

In contrast to what was observed in the first analysis for the Environmental pillar, the significance of these CAARs persisted also under the longest event window, suggesting that in certain sectors, not only the information conveyed by the combined ESG rating is weighted differently, but may also take longer to be fully incorporated into prices.

In conclusion, while this study provides evidence of some market reaction to ESG rating announcements in Europe, particularly for downgrades in the Environmental pillar, the overall impact of this kind of information appears limited in the short term. This suggests that investors may need to consider ESG information in conjunction with other factors when making investment decisions and that the integration of ESG factors into market valuation may be a gradual process.

However, in particular sectors, variations of the ESG score already cause important price movements, suggesting that the relevance of this information depends on the context.

Bibliography

Armitage, S. (1995). ”Event study methods and evidence on their performance”. *Journal of Economic Surveys*, 9(1):25–52.

Berg, F., Kalbel, J. F., and Rigobon, R. (2022). ”Aggregate Confusion: The Divergence of ESG Ratings”. *Review of Finance*, 26(6):1315–1344.

Broadstock, D. C., Chan, K., Cheng, L. T., and Wang, X. (2021). The role of esg performance during times of financial crisis: Evidence from covid-19 in china. *Finance Research Letters*, 38:101716.

Brown, S. J. and Warner, J. B. (1980). ”Measuring security price performance”. *Journal of Financial Economics*, 8(3):205–258.

Brown, S. J. and Warner, J. B. (1985). ”Using daily stock returns: The case of event studies”. *Journal of Financial Economics*, 14(1):3–31.

Chen, Y., Li, T., Zeng, Q., and Zhu, B. (2023). Effect of esg performance on the cost of equity capital: Evidence from china. *International Review of Economics Finance*, 83:348–364.

Cheng, B., Ioannou, I., and Serafeim, G. (2013). ”Corporate social responsibility and access to finance”. *Strategic Management Journal*, 35:1–23.

Corrado, C. J. and Zivney, T. L. (1992). ”The Specification and Power of the Sign Test in Event Study Hypothesis Tests Using Daily Stock Returns”. *The Journal of Financial and Quantitative Analysis*, 27(3):465–478.

Eccles, R. G., Ioannou, I., and Serafeim, G. (2014). ”The Impact of Corporate Sustainability on Organizational Processes and Performance”. *Management Science*, 60(11):2835–2857.

Ederington, L. H. and Goh, J. C. (1998). ”Bond Rating Agencies and Stock Analysts: Who Knows What When?”. *The Journal of Financial and Quantitative Analysis*, 33(4):569–585.

Fama, Eugene F, e. a. (1969). ”The Adjustment of Stock Prices to New Information”. *International Economic Review*, 10(1):1–21.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Fama, E. F. and French, K. R. (1993). ”Common risk factors in the returns on stocks and bonds”. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2014). ”A Five-Factor Asset Pricing Model”. *Fama-Miller Working Paper*.

Glück, M., Hübel, B., and Scholz, H. (2021). ”ESG Rating Events and Stock Market Reactions”. *SSRN*.

Gunnar Friede, T. B. and Bassen, A. (2015). ”ESG and financial performance: aggregated evidence from more than 2000 empirical studies”. *Journal of Sustainable Finance & Investment*, 5(4):210–233.

Hartzmark, S. M. and Sussman, A. B. (2019). ”Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows”. *European Corporate Governance Institute - Finance Working Paper*, 565.

Henderson, G. V. (1990). ”Problems and Solutions in Conducting Event Studies”. *The Journal of Risk and Insurance*, 57(2):282–306.

Kothari, S. and Warner, J. B. (2004). ”The Econometrics of Event Studies”. *SSRN*.

Leite, B. J. and Uysal, V. B. (2023). ”Does ESG matter to investors? ESG scores and the stock price response to new information”. *Global Finance Journal*, 57.

Li, Y., Zhao, Y., Ye, C., Li, X., and Tao, Y. (2024). Esg ratings and the cost of equity capital in china. *Energy Economics*, 136:107685.

MacKinlay, A. C. (1997). ”Event Studies in Economics and Finance”. *Journal of Economic Literature*, 35(1):13–39.

MSCI (2019). ”ESG ratings, measuring a company’s resilience to long-term, financially relevant esg risks”. Accessed: 2024-04-14.

MSCI (2024). ”MSCI Ratings Methodologies”. Accessed: 2024-08-04.

Rzeźnik, A., Hanley, K. W., and Pelizzon, L. (2022). ”Investor Reliance on ESG Ratings and Stock Price Performance”. *SAFE Working Paper No. 310*.

Sassen, R., Hinze, A.-K., and Hardeck, I. (2016). ”Impact of ESG factors on firm risk in Europe”. *Journal of Business Economics*, 86:867–904.

Serafeim, G., Y. A. (2023). ”Stock price reactions to ESG news: the role of ESG ratings and disagreement. *Review of Accounting Studies*, 28:1500–1530.

Welch, K. and Yoon, A. (2020). ”Do high-ability managers choose ESG projects that create shareholder value? Evidence from employee opinions”. *Review of Accounting Studies*.

Whelan, T., Atz, U., Holt, T. V., Clark, C. C., Salazar, P., Liu, Z., and Bruno, C. (2021). ”ESG AND FINANCIAL PERFORMANCE: Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015 – 2020”. *NYU Stern Center for Sustainable Business*.

Acknowledgements

I want to thank my family, my girlfriend Eliana, and anyone who supported me in this beautiful adventure.