How do changes in ESG ratings impact stock returns?

Laura Leirião

Dissertation written under the supervision of
Professor Eva Schliephake

Dissertation submitted in partial fulfilment of requirements for the MSc in Finance, at the Universidade Católica Portuguesa, January 2023.

Abstract

The main purpose of this thesis is to analyse the impact of changes in environmental, social and governance ratings of firms on their stock market returns. This analysis is done for S&P 500 firms, using ESG data from Refinitiv for the 10-year period between 2011 and 2020, by performing an event study where events correspond to the dates when there is a change in ESG, E, S and G scores. These 4 components are analysed separately, because considering aggregate ESG scores alone might have a mitigating effect and not provide such a clear picture of the possible market reactions to changes in each ESG pillar. Events are divided into upgrades, downgrades and unchanged, where upgrades and downgrades are defined as a positive or negative change in the letter rating, respectively, and unchanged rating events represent a small change in scores which does not affect the letter rating. The results from the event study provide evidence that investors react negatively to both upgrades and downgrades in ESG ratings. However, the stock market reaction is stronger in the case of downgrades. There are also some abnormal returns, even though much smaller, prior to the event dates, which can indicate some degree of speculation about the rating changes or information leakage before the ESG scores are updated.

Keywords: ESG, ESG rating, event study, stock returns, abnormal returns

Title: How do changes in ESG ratings impact stock returns?

Author: Laura Isabel Gomes Leirião

Resumo

O principal objetivo desta tese é analisar o impacto que alterações nas classificações de empresas a nível ambiental, social e de governança têm no retorno das ações dessas empresas. Esta análise utiliza dados fornecidos pela plataforma Refinitiv referentes ao período de 2011 a 2020 e foca-se nas empresas incluídas no S&P 500. A metodologia utilizada, denominada *event study,* identifica as datas em que há alterações nestas classificações e avalia a respetiva reação do mercado de ações a estes eventos. Esta avaliação é feita tanto para reações a alterações na classificação geral das empresas (*ESG*), como para alterações em cada uma das 3 componentes individualmente (*E*, *S* e *G*). Um aumento ou diminuição nestas pontuações que resultem, consequentemente, numa alteração da classificação da empresa na escala de D- a A+ são definidos como *upgrades* e *downgrades*, respetivamente. Uma pequena alteração nas pontuações, em qualquer direção, que não resulte na alteração desta letra é definida como *unchanged*. Os resultados deste estudo mostram que os investidores reagem negativamente tanto a alterações positivas como negativas a nível das classificações *ESG, E, S* e *G* das empresas. Contudo, esta reação é mais notável no caso dos *downgrades* e verifica-se não só imediatamente após os eventos, mas também nos dias que os antecedem. Isto sugere que poderá haver algum nível de especulação ou fuga de informação no período que antecede a atualização destas classificações.

Palavras-chave: *ESG*, *ESG rating, event study*, retorno de ações, retornos anormais
Título: Como é que alterações em *ESG ratings* afetam o retorno das ações de empresas?
Autor: Laura Isabel Gomes Leirião

Table of Contents

1. Introduction

The importance of ESG topics has been growing in recent years, which affects both firms and investors, among others. Firms are increasingly reporting on their ESG practices and, consequently, there are more firms being assigned ESG ratings nowadays than there were twenty years ago. For instance, Refinitiv provides ESG ratings dating back to 2002 and currently covers more than 80% of the global market cap, including both public and private firms from developed and emerging countries. Thus, with ESG reporting becoming a more common and widespread practice and a growing number of firms being assigned ESG ratings, it is interesting to understand the extent to which investors are concerned about ESG information and if they take ESG indicators into consideration when making their investment decisions.

Studies have shown that ESG disclosure can have a positive impact on firms' financial performance and attract ESG investors. Meanwhile, as firms are increasingly viewing ESG matters as an opportunity to improve their businesses, investors are seeking to improve their returns. Some authors suggest that ESG investing offers plenty of opportunities for investors to enhance their portfolios' performance and risk-adjusted returns, though others have found no significant improvements when resorting to ESG strategies. To this day it remains up for discussion whether ESG data, and in particular ESG ratings, provide relevant material information to investors, with various research having found contradicting evidence on this matter.

This thesis provides insight into the relationship between firms' ESG performance and their stock market performance. I highlight some of the key challenges associated with ESG topics nowadays which might be driving the lack of consensus among researchers in this field, such as the non-financial nature of ESG information and lack of agreement in ESG ratings from different rating providers. Additionally, I contribute to this debate by performing an event study to analyse how firms' ESG ratings impact their stock returns.

Most of the existing literature on this subject aims to assess if there is a link between firms having higher or lower ESG ratings and their respective stock returns. In addition, this study aims to understand if upgrades and downgrades in ESG ratings signal different information to investors. I therefore distinguish between upgrades and downgrades and estimate their impact on investment decisions, separately.

Furthermore, in my research I differentiate between the environmental, social and governance components of ESG, since aggregating E, S and G scores into one single ESG rating might be misleading. This is particularly important for firms for whose we observe changes of different components with opposite directions because they might offset each other in aggregate.

Regarding the structure of this dissertation, I begin by elaborating on the different findings from existing literature. Then, I provide a summary of the data used and the most relevant descriptive statistics. This is followed by a description of the methodology applied and the respective results from this event study. Lastly, I present the main limitations of this work and summarize the key takeaways and results.

2. Literature Review

The existing literature surrounding ESG topics has been considerably growing in the past years. Yet, there remains a lot of controversy regarding the impact that ESG data can have on investors' decisions and stock performance. This is quite notorious for instance in studies analysing market reactions to ESG news, with different authors presenting substantially distinct findings.

In Krüger's (2015) analysis of stock market reactions to events related to CSR, he finds that there is a strong negative response to negative events and an also negative reaction, even though weaker, to positive events. However, he adds that in cases when positive news are the result of an attempt to compensate for firms' previous negative behaviour, it is possible to see an increase in firms' average stock prices.

Capelle-Blancard and Petit (2019) also investigate the impact of positive and negative ESG news on the stock market, covering both ordinary and extreme events, but find mostly significant results in the case of negative announcements, which result in an average decrease of 0.1% on firms' market value. Additionally, they state that this negative reaction is lessened when the sector in which the firm operates has a good ESG reputation.

Conversely, Serafeim and Yoon (2022) have found that the market reaction is stronger for positive news than negative ones, showing that positive unexpected ESG news increase stock prices. They also find that the investors are very selective in the sense that their reactions vary according to the type of news. In particular, financially fundamental information appears to be the most impactful.

Even though the direction of the market reactions is not yet clear, these findings suggest that investors are concerned, at least to some extent, with matters related to ESG and that information of this nature is relevant to their investment decisions.

There is also extensive literature investigating the relationship between ESG performance and stock performance and, similarly, the results are heterogeneous. However, research dedicated to the analysis of stock market reactions specifically to ESG rating changes, as opposed to ESG rating levels, is much more limited.

Shanaev and Ghimire (2022) investigate the impact of ESG rating upgrades and downgrades on stock performance and find that upgrades are responsible for an overall increase in stock prices, while downgrades seem to have the opposite effect. They also provide evidence that the magnitude of the stock market reaction is considerably stronger for downgrades compared to upgrades. Furthermore, they assess whether previous ESG rating levels have any influence on the size of the stock market reaction to these changes and find that investors react more strongly to rating upgrades and downgrades in cases of initially high-rated companies (ESG leaders) compared to initially low-rated ones (ESG laggards).

Similarly, Glück, Hübel and Scholz (2022) perform an event study to assess whether upgrades and downgrades in ESG scores generate abnormal returns, but they find significant results only in the case of downgrades and, more specifically, in the environmental and social components of the ESG rating. Additionally, they also test the hypothesis that abnormal returns from changes in ESG ratings are dependent on firms' previous ESG scores, but ultimately find no evidence of this.

Overall, there is a noticeable lack of consensus in previous academic research surrounding the topic of ESG. Part of the reason why this is the case might stem from the fact that the ESG data being used in different studies is often provided by different sources. Billio, Costola, Hristova and Latino (2021) show that there is a discrepancy in the rating criteria used by distinct rating agencies and that these lack a uniform and standardized definition of ESG and its components, which might cause them to evaluate the same companies differently.

There is also a significant amount of research aimed at identifying specific factors which are driving such divergences in ESG ratings. Brandon, Krueger and Schmidt (2021) identify three main ones: there seems to be less disagreement between ratings of more profitable firms, and more disagreement between ratings of larger firms as well as firms which do not have a credit rating.

Additionally, Berg, Koelbel and Rigobon (2019) provide evidence of a "rater effect", which indicates that the assessment of a firm's performance in one category is influenced by how well the firm scored in other categories. In other words, if one firm is rated highly in one dimension, then the same rating agency is more likely to rate that same firm highly as well in other dimensions. This effect suggests that ESG ratings are somewhat biased and that investors should be particularly mindful of the underlying data.

Another critical aspect that should be taken into consideration when dealing with ESG data is the extent of its reliability and accuracy, as it is probable that there is some amount of bias as well in ESG reporting. Jonsdottir, Sigurjonsson, Johannsdottir and Wendt (2022) highlight that companies have a tendency towards "window dressing" when reporting on their ESG practices, emphasizing mostly their positive behaviours. This often serves as a reasonable explanation as to why some researchers find stronger market reactions to negative ESG events than to positive ESG events, since negative news and rating downgrades are more likely to reflect new information which companies are typically less eager to disclose. Additionally, Giudice and Rigamonti (2020) state that since non-financial reporting is not as developed as financing reporting yet and lacks standardized practices, it is harder to ensure the accuracy of self-reported ESG information, which in turn affects the precision of ESG scores.

Overall, the aforementioned findings from existing ESG-themed literature are pertinent to this thesis since they suggest that ESG data is indeed relevant to investors' decision-making processes, even though the magnitude of this relationship is not yet clear. In this study, I intend to assess how investors react to events of both positive and negative nature, i.e., to ESG rating upgrades and downgrades. Additionally, previous research suggests that ESG data is likely to carry some bias, which is in part due to the lack of standardized reporting methods which affects the accuracy of firms' ESG reports. The inconsistency of rating criteria and metrics used by rating agencies in the computation of ESG scores, and consequential disagreement in ratings, also contributes to exacerbate the reliability issue of ESG information. These factors must then be taken into consideration when interpreting the results of a study which uses ESG data, as is the case of this thesis.

3. Data & Descriptive Statistics

This study uses ESG data of S&P 500 constituents, between 2011 and 2020. The data was retrieved from Refinitiv Datastream, one of the most complete databases in terms of ESG data.

Refinitiv's ESG scores are based on company-reported data and are typically updated on a yearly basis, with a few exceptions when companies undergo considerable changes in their reporting practices, in which cases there might be additional updates.

The methodology applied by Refinitiv to compute these scores covers over 630 different metrics, of which around 180 are selected for each industry, according to their relevance to that industry, to be used in the calculation of the ESG scores. In turn, these are then used to compute separate scores for each one of the three ESG pillars and, lastly, the E, S and G scores of each company are aggregated into an overall ESG rating. The ESG scores range between 0 and 100, corresponding to a letter rating from D- (ESG laggards) to A+ (ESG leaders), similarly to the terminology employed for credit ratings.

For this study, it is pertinent to use data not only regarding firms' overall scores but also for the environmental, social and governance components separately. The distinction is important because considering only the overall ESG performance of firms might have a mitigating effect on the results whenever firms' performance simultaneously improves in one aspect and deteriorates in another. Additionally, the separation also allows to observe whether investors react differently to changes in different components.

As a next step, to perform an event study it is necessary to identify the event days, which in this thesis correspond to the dates when there is a change in companies' ESG, E, S and G scores. At this point, no differentiation is made yet regarding the magnitude of each change.

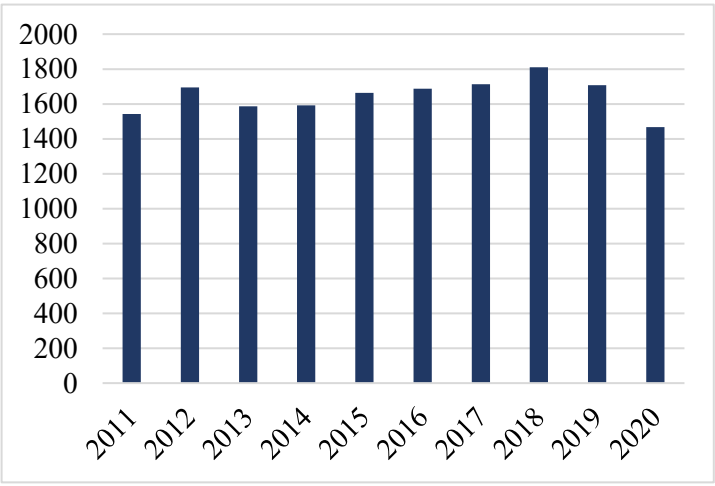
The goal here is to analyse the stock market reactions to these events, which implies that event days must be trading days so that firms' stock returns are observable at each point in time. However, in this sample this is not always the case, with a significant number of events coinciding on January 1st every year. For this reason, some adjustments are made in such circumstances, by moving the event date to the closest trading day after the actual event took place. Consequently, this implies that even though all event windows cover the same number of trading days, some might have up to two more calendar days than others.

Additionally, since Refinitiv provides ESG scores for both public and private companies, some of the ESG data collected concerns a period preceding some companies' IPOs. Hence, events registered before the company became public are excluded from this study, since they are irrelevant as the company was not yet listed in the stock market and thus had no stock market returns.

Furthermore, to distinguish events based on their magnitude and direction of the rating change, they are divided into upgrades, downgrades and unchanged ratings. An upgrade occurs when there is a positive change in the ESG score which results in a changed letter rating, i.e., the rating improves by at least one letter. Likewise, a downgrade happens when there is a negative change in the ESG score which also causes the letter rating to change, in this case by decreasing at least one letter. Lastly, an unchanged rating is defined as an insignificant change in ESG scores which does not lead to a change in the letter rating. These events technically cover both small increases as well as small decreases in ESG scores, as long as the overall rating remains unchanged.

Additionally, performing an event study requires choosing a model for estimating expected returns. In this thesis, I apply three different models: the 1-factor model, the 6-factor model and the market model. Consequently, it is necessary to collect data for all these factors: market excess return (R_M-R_f), size (SMB), value (HML), profitability (RMW), investment (CMA) and momentum factors (WML). This data was retrieved from the Kenneth R. French data library. I elaborate on this further in the Methodology section of this work.

In total, there are 16466 events registered between 2011 and the end of 2020, including upgrades, downgrades, and unchanged ratings in the 3 ESG pillars as well as in overall ESG scores. Events are fairly distributed throughout the sample period, as depicted in Graph 1, with an average of 1645 events recorded each year.

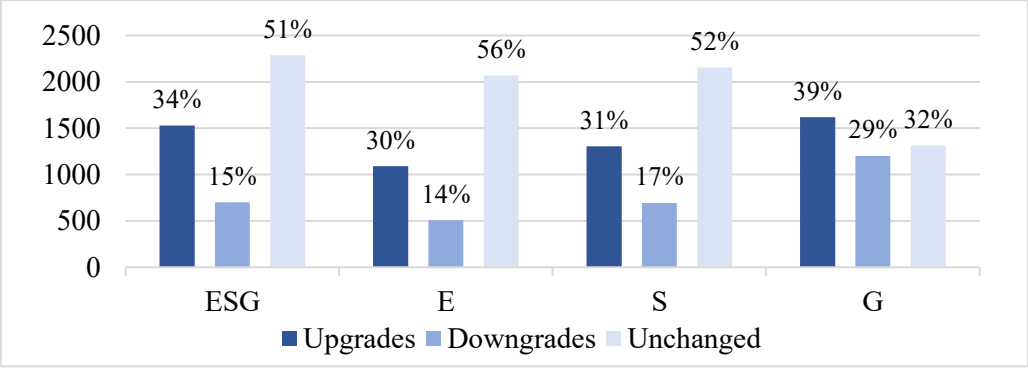


Graph 1: Distribution of events between 2011 and 2020.

Analysing the distribution of rating upgrades, downgrades and unchanged ratings across the four ESG, E, S and G dimensions, it is possible to observe that the total amount of upgrades exceeds the number of downgrades in all categories. Also, a significant portion of the events

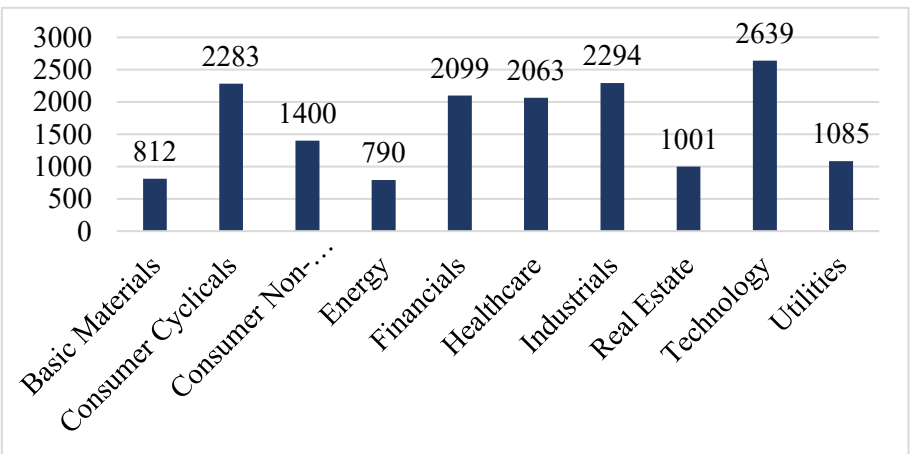
do not represent any change in the letter rating of the ESG pillar, with a total of 7818 unchanged rating events recorded in this period (approximately 47% of all the events).

Looking more closely at each dimension, G scores seem to be the less steady, with 68% of the events related to governance scores representing either a rating upgrade or downgrade, whereas for all the remaining categories this figure is below 50%, as shown in Graph 2.



Graph 2: Distribution of changes in ratings for ESG, E, S and G dimensions.

Regarding the distribution of rating events per industry, it is in line with the number of firms in the sample from each sector, i.e., the percentage of events relating to each sector is approximately proportionate to the weight of firms from that sector on the S&P 500. Graph 3 shows the total number of events registered in each industry in the sample period. Almost 70% of the events are related to the financial, healthcare, industrials, technology, and consumer cyclicals sectors. For the identification of the industry, the "*The Refinitiv Business Classification*" (TRBC) is used here, since this is also the one which Refinitiv applies in its calculation of ESG scores.



Graph 3: Distribution of total events(ESG, E, S and G) per industry.

Moreover, since upgrades, downgrades and unchanged ratings are measured by the change in ESG letter ratings, it is useful to convert the scores presented on a scale from D- to A+ into a

scale from 1 to 12. For example, scores between 0 and 8.33 which correspond to a D- rating are converted to 1, and so on. This allows for easier computations as well as a more intuitive interpretation of the data's descriptive statistics.

Table 1 provides descriptive statistics regarding the ESG data collected as well as the changes in ESG scores. In all 4 categories (ESG, E, S and G) there are extreme cases of firms which register the lowest possible letter rating, D-, as well as firms which are attributed the best letter rating, A+. Overall, the average scores are quite similar across all ESG pillars: around 7,5 for social and governance scores and slightly lower for environmental scores which average 6,8, with the aggregate ESG scores averaging 7,3. These averages correspond to B- and B scores on the letter scale.

In terms of changes in ESG scores, there are downgrades and upgrades of substantial magnitude throughout the sample period, with some firms' E, S, and G ratings deteriorating by 8 letters at once. Conversely, there are other events when firms' scores are updated and these improve by up to 7 letters.

	ESG	Change in ESG	E	Change in E	S	Change in S	G	Change in G
Minimum	1	-9	1	-8	1	-8	1	-8
Maximum	12	7	12	7	12	7	12	7
1st Quartile	6	0	4	0	6	0	6	-1
Median	8	0	7	0	8	0	8	0
3rd Quartile	9	1	9	1	9	1	10	1
Mean	7,330	0,258	6,844	0,301	7,547	0,272	7,597	0,185
Standard Deviation	2,206	0,888	3,022	1,141	2,476	1,025	2,490	1,572
Kurtosis	2,512	7,582	2,043	8,774	2,283	7,242	2,451	4,280
Skewness	0,390	0,355	0,336	1,053	0,212	0,953	0,396	0,024
Observations	4518	4518	3665	3665	4154	4154	4129	4129

Table 1: Descriptive statistics for both ESG, E, S and G scores, and changes in these scores.

Moreover, Table 2 presents some descriptive statistics of upgrades and downgrades in each ESG component and in the aggregate scores. It is possible to observe that the magnitude of rating changes is, on average, larger for upgrades than for downgrades across all four categories. However, when analysing the standard deviations, it is also clear that the size of upgrades is more dispersed than the size of downgrades. Additionally, in absolute value terms, downgrades register the largest changes in every dimension during this period.

It is also relevant to emphasize that the minimum values of upgrades and downgrades shown in this table are very small, which is a consequence of using a definition based on letter ratings instead of changes of a given size or percentage changes.

	ESG	E	S	G
Upgrades				
Minimum	0,39	0,06	0,24	0,37
Maximum	58,18	61,53	58,56	62,05
1st Quartile	4,66	4,81	4,87	6,75
Median	7,34	8,55	8,22	10,70
3rd Quartile	11,13	15,94	13,49	16,78
Mean	8,55	11,78	10,27	12,88
Standard Deviation	5,70	9,93	7,73	8,77
Observations	1530	1090	1306	1619
Downgrades				
Minimum	0,25	0,09	0,20	0,26
Maximum	69,96	71,95	67,93	70,24
1st Quartile	3,37	2,84	3,20	6,34
Median	5,29	5,80	5,26	10,52
3rd Quartile	7,78	9,55	8,38	16,38
Mean	5,93	7,84	6,31	12,44
Standard Deviation	4,25	7,62	4,82	8,39
Observations	700	508	695	1200

Table 2: Descriptive statistics of upgrades and downgrades in ESG, E, S and G scores, in absolute terms, on a scale from 1-100.

Furthermore, defining an upgrade/downgrade as a change in the letter rating results in some rating changes of a given size (on a scale from 1 to 100) being categorized as an upgrade or a downgrade, while other changes of the same size are allocated to the "unchanged" category. Hence, it is relevant to look at the descriptive statistics of events classified as unchanged to better understand the composition of this category. Additionally, it is helpful to distinguish between "unchanged" events where the ESG score increases (+) and events where it decreases (-), since this will provide some insight about the distribution of events between upgrades and unchanged ratings, as well as between downgrades and unchanged ratings. This is shown in Table 3.

	ESG		E		S		G	
	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)
Minimum	0,01	-8,05	0,01	-7,29	0,01	-8,08	0,01	-8,13
Maximum	7,95	-0,01	8,33	-0,01	8,10	-0,01	8,22	-0,01
1st Quartile	0,88	-3,07	0,61	-2,32	0,67	-2,66	1,00	-3,56
Median	1,95	-1,74	1,51	-1,23	1,57	-1,48	2,27	-1,82
3rd Quartile	3,35	-0,82	2,96	-0,56	3,04	-0,69	3,92	-0,80
Mean	2,31	-2,12	2,00	-1,65	2,02	-1,87	2,65	-2,30
Standard Deviation	1,76	1,61	1,71	1,44	1,72	1,50	1,94	1,83
Observations	1267	1021	1154	913	1095	1058	692	618
Total Events	2288		2067		2153		1310	

Table 3: Descriptive statistics of unchanged rating events in the ESG, E, S and G categories, for small positive (+) and negative (-) changes separately.

Even though an unchanged rating event means that the ESG letter rating of a company is unaffected when Refinitiv updates its ESG scores, there is still some degree of fluctuation in the actual scores on a scale from 1 to 100. Naturally, the maximum and minimum values for this variation, which are presented in the table above, cannot exceed the 8,33 threshold in absolute terms, since this value corresponds to the range covered by each letter on the 12-letter scale from D- to A+.

The most relevant information provided by Table 3 are the mean values. The first column of each ESG component shows that in cases when the letter rating does not change but the score increases, this increase is on average close to 2. Similarly, in the case of events where the score decreases slightly but the letter rating remains unchanged, this decrease is on average also close to 2, in absolute terms, or even smaller. However, looking back at the average values presented on Table 2, the average variation in ESG, E, S and G scores for both upgrades and downgrades is much more significant: always above 8 for upgrades and above 5 for downgrades.

Thus, even though the definition of upgrades, downgrades and unchanged ratings applied here has some constraints, it still seems adequate for the purpose of this study because the average change in ESG ratings covered by events labelled "unchanged" is substantially smaller when compared to the respective averages of upgrades and downgrades.

4. Methodology

The use of the event study methodology in academic research related to ESG and CSR topics is quite common. In their research, Kruger (2015) and Capelle-Blancard and Petit (2019) perform event studies to examine the market reactions to positive and negative events related to CSR and ESG, respectively. Both papers defend that one of the key strengths in resorting to this methodology rests on its ability to mitigate reverse causality problems. More recently, Glück, Hübel and Scholz (2022) also opt to carry out an event study to investigate stock market reactions to changes in firms' ESG ratings. This piece of research particularly influenced and served as a frame of reference for the analysis that I perform throughout this thesis.

Furthermore, event studies rely heavily on the assumption of efficient markets, i.e., the hypothesis that markets will rapidly adjust to new information that arises. The rationale behind this methodology is that this new information, which comes to light when certain events take place, is expected to be reflected on the stock market shortly after. Hence, looking at the period surrounding these events, it is possible to assess whether investors react positively, negatively,

or not at all, by comparing how stock returns in the short-term deviate from what would normally be expected, had the event not occurred. Thus, the vast applications of this methodology seem to be in line with the purpose of this work, so I choose to conduct an event study to analyse the impact of changes in ESG scores on firms' stock returns.

Firstly, I begin by defining an event window and an estimation window surrounding each of the 16466 rating events. The event window necessarily includes the event day itself and, even though event windows in different studies often vary in length, it is common for them to cover also a period preceding the event and another period after the event. Glück, Hübel and Scholz (2022) suggest that the former period is useful to account for the possibility of there being any leakage of information before the actual event takes place, while the latter is concerned with the short-term abnormal returns caused by the event.

In this case, the event window is defined as the period starting 5 trading days before and ending 10 trading days after the event: [-5, 10]. As previously mentioned in the data section of this work, since it is necessary for every day inside both the estimation and event windows to be trading days, the event day is often pushed forward to the closest trading day following the actual event. Consequently, even though all the event windows technically consist of 16 trading days, they might in reality cover one or two additional days in cases where adjustments were made.

Regarding the estimation window, there is also quite a lot of flexibility in the decision of the number of days it should contain. However, usually it does not include days which are part of the event period since this could induce some bias in the calculation of the expected returns. In this study, rating events mostly have an annual frequency, so it is possible to use an estimation window of 200 trading days without having different estimation windows overlapping. This window begins 205 trading days before the event day and ends 1 day before the event window starts: [-205, -6]. For the sake of robustness, an estimation window of 245 trading days is also used, which approximately covers one calendar year prior to the event.

After defining both the event and estimation windows, it is possible to compute abnormal returns (AR) for each day inside event windows. Abnormal returns represent how much the realized return on a firm's stock on a certain day differs from its expected return for that day, and are computed using either one of the two formulas below:

$$AR_{i,t} = R_{i,t} - E(R_{i,t}) \quad (1)$$

$$AR_{i,t} = ER_{i,t} - E(ER_{i,t}) \quad (2)$$

$AR_{i,t}$ is the abnormal return of stock i at time t, $R_{i,t}$ is the realized return of stock i at time t, $E(R_{i,t})$ is the expected return of stock i at time t, $ER_{i,t}$ is the realized excess return of stock i at time t and $E(ER_{i,t})$ is expected excess return of stock i at time t.

The calculations of realized returns and realized excess returns are very straightforward as they can be observed in the market and do not have to be inferred. I retrieved from Datastream the return index for the S&P 500 firms between 2010 and 2020 with a daily frequency and used this to compute the daily stock returns for each company. Here, the daily stock returns from 2010 are necessary, since the estimation windows for events taking place throughout 2011 will cover days from the previous calendar year.

On the other hand, expected returns are not visible in the market since they represent what the stock return on a certain day would have been if the event had not taken place, hence they must be estimated. As previously stated in the Data section of this work, I use three different models to calculate expected (excess) returns: the 1-Factor model (CAPM), the 6-Factor model and the Market Model (MM). Abnormal returns are calculated using the first equation (1) in cases where the expected returns are estimated using the Market Model (MM), and they are calculated using the second equation (2) when expected excess returns are calculated using either the 1-factor model (CAPM) or the 6-factor model.

Applying the 1-factor model (CAPM), expected excess returns for each day within an event window, $E(ER_{i,t})$, are estimated as follows:

$$\text{1-Factor Model: } E(ER_{i,t}) = \alpha_i + \beta_i [R_M - R_f] \tag{3}$$

To obtain the coefficient (β_i), realized excess returns are regressed on the market risk premium, for the period covered by the estimation window. The market risk premium (MRP) is defined as the difference between the US market return and the risk-free rate on any given day.

Similarly, using the 6-Factor model to compute expected excess returns follows the same reasoning, but with the addition of five more factors to the regression equation, as proposed by Fama and French: size (SMB), value (HML), profitability (RMW), investment (CMA) and momentum (WML). In this case, regressing realized excess returns on the 6 factors across the estimation window generates 6 coefficients, which are then used to compute expected returns, following the expression below:

$$\begin{aligned}\text{6-Factor Model: } E(ER_{i,t}) = \alpha_i + \beta_i [R_M - R_f] + \beta^{SMB}{}_i\, SMB_t + \beta^{HML}{}_i\, HML_t \\ + \beta^{RMW}{}_i\, RMW_t + \beta^{CMA}{}_i\, CMA_t + \beta^{WML}{}_i\, WML_t\end{aligned} \tag{4}$$

Lastly, applying the market model in this context consists of regressing daily stock returns over the estimation window on the US market return to obtain the coefficient (β_i) and then compute expected returns as follows:

$$\text{Market Model: } E(R_{i,t}) = \alpha_i + \beta_i R_M \quad (5)$$

As a next step, I divide the event window into 2 periods, [-5,-1] and [0,10], to better assess whether there are any market reactions surrounding changes in ESG scores and if they occur before or after the event. Then, for each one of the ESG categories (ESG, E, S and G scores), I split rating events into groups, according to whether they are upgrades, downgrades or unchanged ratings. For each of these 12 sub sections created, I compute the cumulative abnormal returns (CAR) of each event, for the periods [-5,-1] and [0,10] separately. This is done by summing the daily abnormal returns within each of these two periods, as follows:

$$CARi(t_1, t_2) = \sum_{t=t1}^{t2} AR_{i,t} \quad (6)$$

Lastly, cumulative average abnormal returns (CAAR) are calculated by adding the CAR of all the events in the respective subsection, to obtain CAAR[-5,-1] and CAAR[0,10], as follows:

$$CAAR(t_1, t_2) = \frac{1}{N} \sum_{i=1}^{N} CARi(t_1, t_2) \quad (7)$$

Where t_1 is the first day of each period of the event window and t_2 is the last day of each period of the event window; i corresponds to each event of every subsection and N corresponds to the total number of events in each subsection.

To test the significance of the cumulative average abnormal returns obtained from the event study, I apply a parametric cross-sectional test where the test statistic is given by the formula:

$$t_{CAAR} = \frac{CAAR\sqrt{N}}{s} \quad (8)$$

The denominator, *s,* corresponds to the standard deviation of the CAR, defined as the square root of the variance given by:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (CAR_i - CAAR)^2 \quad (9)$$

5. Results

Performing an event study is helpful to understand if certain events have any impact on firms' stock market returns, which can be determined by assessing whether events generate abnormal returns. Here, the goal is to understand how upgrades and downgrades in ESG, E, S and G scores influence the stock market, so it is important to look at the cumulative average abnormal returns (CAAR) for each of these dimensions. These are presented separately for upgrades and downgrades so that it is possible to differentiate between investors' reactions to events of a positive and negative nature. Lastly, it is necessary to test whether CAAR are significantly different from 0. If they are either positive or negative and significantly different from 0, then this constitutes evidence supporting the hypothesis that there are market reactions to changes in ESG ratings. Conversely, if the results from the event study are not statistically significant, then it is not possible to make any conclusions in this regard. Table 4.1 summarizes the main results obtained from this study.

Regarding the events associated with a rating downgrade, the event study finds that they generate negative CAAR in the 5-day period prior to the event day, significant at the 1% or 5% level, in all dimensions except for E scores. This is still observable when using each of the three models for estimating expected returns, with the 1-factor and 6-factor models presenting the highest values for CAAR, reaching -0,30% for ESG rating downgrades. Regarding the period after (and including) the event, the CAAR[0,10] from rating downgrades are significant only for the environmental and social dimensions when using the 1-factor and the market model, and significant across all four dimensions when using the 6-factor model. The CAAR[0,10] varies from -0,70% (Market Model) to -0,75% (1-factor) for downgrades in E scores, and from -0,55% (MM) to -0,75% (6-factor) for downgrades in S scores. These results are in line with plenty of the existing research on this topic, which supports that investors react negatively to negative events, including not only ESG rating downgrades but also other types of ESG news of a negative nature. However, the magnitude of this reaction is still highly debated. In this study, the absolute values of the abnormal returns following rating downgrades suggest that investors react strongly to these events and, thus, the latter are of economic significance.

In terms of the market reactions to rating upgrades, the 1-factor and 6-factor models find significantly negative CAAR for all four ESG, E, S and G dimensions in the five days prior to the event. In most cases however, the CAAR[-5,-1] for upgrades are smaller than for downgrades. After the event day, all three models find significantly negative CAAR for upgrades in ESG and G scores. The 6-factor model also shows negative CAAR[0,10] in E

scores at the 1% significance level and in S scores at the 10% significance level. These findings are particularly surprising since they contrast with the results from most of similar previous studies, which largely find either a positive reaction to positive ESG news or no reaction at all. Nevertheless, there is some literature which provides evidence consistent with these results. In particular, Krüger (2015) shows that, in general, investors react negatively to positive news and that this is more evident in cases where firms are likely to have agency problems. Additionally, he states that there is a stronger tendency for large firms to suffer from agency problems. Since this study uses data from S&P 500 companies, which are substantially large firms in terms of market capitalization, this seems to be a plausible explanation as to why the results show significantly negative abnormal returns from ESG rating upgrades.

Overall, the magnitude of cumulative average abnormal returns caused by rating upgrades is considerably smaller than that of the CAAR of downgrades in most cases. For example, the 1-factor and 6-factor models find CAAR[-5,-1] of -0,10% and -0,11%, respectively, for ESG rating upgrades while the same models find CAAR[-5,-1] of -0,30% for ESG rating downgrades. The is also observable in the cumulative average abnormal returns for the period after, and including, the event day. All three models find larger CAAR[0,10] for rating downgrades than upgrades in ESG, E and S scores. However, investors' reactions to upgrades in G scores in the days following the events are stronger than their reactions to downgrades in the same scores. These findings are, once more, mostly consistent with Krüger (2015), who finds a stronger reaction to negative events compared to positive events.

Additionally, the results from the event study show significant negative CAAR[-5,-1] and CAAR[0,10] for unchanged rating events when using the 1-factor and 6-factor models, for all ESG dimensions. To understand these findings, it is important to highlight that the definition of an unchanged rating used in this study encompasses modest changes in ESG scores, both increases and decreases in ESG, E, S and G scores, which are small enough that the letter rating remains unaffected. Hence, it is necessary to consider that investors might be sensitive even to the smallest change in ESG scores, in which case the abnormal returns here attributed to unchanged rating events might instead be reactions to small upgrades or downgrades. This shows that definition of the events is crucial to the interpretation of the results, which I discuss further in the Limitations section of this paper. Alternatively, it is also possible that these negative abnormal returns result from investors' anticipating increases in ESG scores which do not materialize.

CAAR[-5,-1]	ESG scores	E scores	S scores	G scores
1-Factor Model				
Upgrade	-0,0010*	-0,0015**	-0,0021***	-0,0012**
	(-1,82)	(-2,18)	(-3,51)	(-2,34)
Downgrade	-0,0030***	-0,0013	-0,0021**	-0,0026***
	(-4,04)	(-1,32)	(-2,51)	(-4,27)
Unchanged	-0,0020***	-0,0020***	-0,0017***	-0,0018***
	(-4,45)	(-4,36)	(-3,81)	(-3,00)
6-Factor Model				
Upgrade	-0,0011**	-0,0021***	-0,0022***	-0,0016***
	(-2,03)	(-3,08)	(-3,78)	(-2,99)
Downgrade	-0,0030***	-0,0014	-0,0024***	-0,0026***
	(-4,27)	(-1,44)	(-2,97)	(-4,34)
Unchanged	-0,0022***	-0,0019***	-0,0018***	-0,0018***
	(-5,00)	(-4,40)	(-4,23)	(-3,11)
Market Model				
Upgrade	-0,0009	-0,0012*	-0,0016***	-0,0010**
	(-1,54)	(-1,80)	(-2,79)	(-1,99)
Downgrade	-0,0028***	-0,0010	-0,0019**	-0,0023***
	(-3,79)	(-1,02)	(-2,27)	(-3,83)
Unchanged	-0,0017***	-0,0017***	-0,0014***	-0,0015***
	(-3,87)	(-3,86)	(-3,29)	(-2,59)
CAAR[0,10]	**ESG scores**	**E scores**	**S scores**	**G scores**
1-Factor Model				
Upgrade	-0,0029**	-0,0032**	-0,0003	-0,0032**
	(-2,16)	(-2,10)	(-0,22)	(-2,52)
Downgrade	-0,0037	-0,0075**	-0,0064***	-0,0023
	(-1,57)	(-2,54)	(-2,67)	(-1,37)
Unchanged	-0,0030***	-0,0021**	-0,0033***	-0,0031**
	(-2,74)	(-1,97)	(-3,01)	(-2,14)
6-Factor Model				
Upgrade	-0,0043***	-0,0061***	-0,0027*	-0,0044***
	(-3,30)	(-4,06)	(-1,93)	(-3,49)
Downgrade	-0,0050**	-0,0071**	-0,0075***	-0,0042**
	(-2,10)	(-2,38)	(-3,13)	(-2,47)
Unchanged	-0,0052***	-0,0038***	-0,0048***	-0,0053***
	(-4,79)	(-3,51)	(-4,43)	(-3,74)
Market Model				
Upgrade	-0,0022*	-0,0021	0,0002	-0,0027**
	(-1,66)	(-1,37)	(0,11)	(-2,11)
Downgrade	-0,0031	-0,0070**	-0,0055**	-0,0014
	(-1,32)	(-2,37)	(-2,33)	(-0,84)
Unchanged	-0,0022**	-0,0013	-0,0026**	-0,0022
	(-2,02)	(-1,25)	(-2,34)	(-1,58)

Table 4.1[1]*: Cumulative Average Abnormal Returns of rating upgrades, downgrades and unchanged ratings, with respect to ESG, E, S and G scores, for the 5-day period before the*

[1] *, ** and *** indicate that the CAAR are statistically significant at the 10%, 5% and 1% level, respectively, and the values of the t-statistic are shown in parenthesis.

event day (CAAR[-5,-1]) and for the 11-day period after and including the event day itself (CAAR[0,10]). These results are computed using 3 different models for estimating expected returns, and an estimation window of 200 trading days.

For robustness purposes, I additionally apply the same methodology using an estimation window of 245 trading days. The results are presented in Table 4.2 in the Appendix section. Overall, they are very similar to the ones obtained with an estimation window of 200 trading days.

As a next step, I extend the analysis of the results from the event study to understand whether investors react differently to rating changes when firms have higher or lower ESG, E, S and G scores before the event. Previous research has focused on this matter, but the findings are dispersed. Shanaev and Ghimire (2022) find evidence that there is a stronger market reaction to rating changes when firms have initially higher scores than when they have initially lower scores. Glück, Hübel and Scholz (2022) also test for this, however they do not find any evidence of firms' previous rating levels influencing investors' reactions. Thus, it is interesting to assess whether there is any sign of abnormal returns in this sample being somewhat driven by the previous ESG levels.

To do this, I differentiate between firms which are leaders and those which are laggards according to their ratings. Firms which have an ESG rating equal to or higher than A- are defined as ESG leaders, while firms with an ESG rating equal to or lower than D+ are defined as ESG laggards, and the same thing for all the other categories. Then, I compute CAAR for leaders and laggards separately, for upgrades, downgrades, and unchanged rating events in each category. This process is similar to the computation of the cumulative average abnormal returns presented in Tables 4.1 and 4.2, except in this case we are only interested in the events when firms had previous ratings below D+ or above A-. Here, I use the 1-factor model (CAPM) in the estimation of expected excess returns.

Table 5 and Table 6 show the cumulative average abnormal returns of upgrades, downgrades and unchanged ratings for leaders and laggards separately, for the period prior to the event date and for the period after, and including, the event date, respectively.

	ESG	E	S	G
Upgrades CAAR[-5,-1]				
Leaders	-0,0031	-0,0014	-0,0003	-0,0022
Laggards	-0,0031	-0,0043***	-0,0057***	-0,0021
Downgrades CAAR[-5,-1]				
Leaders	-0,0015	-0,0003	0,0005	-0,0011
Laggards	-0,0031	-0,0008	-0,0052	-0,0017
Unchanged CAAR[-5,-1]				
Leaders	-0,0013	-0,0011	-0,0022**	-0,0009
Laggards	-0,0030*	-0,0015*	-0,0014	-0,0045***

Table 5: CAAR[-5,-1] of rating upgrades, downgrades and unchanged ratings of leaders and laggards, using the 1-factor model in the estimation of expected excess returns.

	ESG	E	S	G
Upgrades CAAR[0,10]				
Leaders	-0,0084	-0,0092*	0,0021	-0,0029
Laggards	-0,0005	-0,0053*	-0,0036	-0,0045
Downgrades CAAR[0,10]				
Leaders	-0,0075	-0,0091	-0,0075	-0,0067*
Laggards	0,0420	-0,0086	0,0025	0,0076
Unchanged CAAR[0,10]				
Leaders	-0,0008	-0,0013	-0,0031	-0,0021
Laggards	-0,0079	-0,0040*	-0,0030	0,0031

Table 6: CAAR[0,10] of rating upgrades, downgrades and unchanged ratings of leaders and laggards, using the 1-factor model in the estimation of expected excess returns.

In the period preceding the event, there are no significant CAAR[-5,-1] from downgrades in ratings of leaders and laggards for any of the ESG components. However, there are statistically significant abnormal returns from upgrades in E and S scores, -0,43% and -0,57%, respectively, when calculating the CAAR[-5,-1] for laggards. These values are considerably higher than the CAAR[-5,-1] from upgrades in E and S scores shown in Table 4.1, 0,15% and -0,21% respectively, which include all the events corresponding to rating upgrades in the sample. This suggests that investors, on average, react more strongly to rating upgrades when firms have previously lower environmental and social scores. It is not possible to make any conclusions.

In the period after and including the event, there are also very few statistically significant cumulative average abnormal returns. It is only worth highlighting the CAAR[0,10] from upgrades in E scores both for leaders and laggards of -0,92% and -0,53%, respectively, significant at the 10% level. This implies that, on average, investors seem to react more strongly to upgrades in environmental scores when firms have previously higher ratings in this category than when they have previously lower ones.

Overall, the results from Tables 5 and 6 do not seem enough to draw definitive conclusions regarding the influence of previous ESG, E, S and G levels on investors reactions to rating changes, since most of the results are not significant. Moreover, it should be noted that this step of the analysis is constrained by the fact that computing CAAR for leaders and laggards only implies that the sample will cover a smaller number of events.

6. Limitations

To ensure the robustness of the results from this study, I use 3 different models in the estimation of expected returns, as well as estimation windows of different sizes. Additionally, I test the significance of all the results obtained for the CAAR at the 1%, 5% and 10% levels. Nevertheless, it is natural that there are still some limitations to this study that must be considered when contemplating the results presented.

As previously discussed in the literature review section of this paper, the quality and accuracy of ESG data is frequently questioned. There is a substantial lack of consensus in terms of ESG ratings provided by different rating agencies, which implies that collecting ESG data from different sources might lead to different results. This study could benefit from using ESG data from more than one database to help attenuate existing biases in ESG scores.

Moreover, the definition adopted to divide rating events into categories is a choice that strongly influences the results obtained from the event study. Since upgrades, downgrades and unchanged rating events are defined based on letter scores, the size of the actual change in the ESG score can vary considerably. Consequently, some events categorized as rating upgrades can correspond to a smaller change in the ESG score, on a scale from 1-100, than other events which are categorized as unchanged because their letter rating does not change, even though they correspond to a larger rating increase on the 1-100 scale. On average, however, the percentage change in ESG, E, S and G ratings of unchanged rating events are around 2 (as shown in Table 3), which is much smaller than the average change of both upgrades and downgrades in ratings of the same categories (as shown in Table 2). This suggests that even though some events defined as “unchanged” might be more relevant than some upgrades and downgrades, the unchanged events overall cover mostly small rating changes and, thus, the constraints of using this definition are less critical. On the other hand, an implication of this is that investors might be sensitive even to the smallest increase or decrease in E,S and G scores, which here is not always captured by the event study, since some of these events are classified

as unchanged ratings. This helps explain why the results show significant abnormal returns surrounding events which, in theory, should not have any impact on stock markets.

Lastly, it is worth highlighting that the results from this study contain some bias due to what is often referred to as the "clustering effect". Event dates for ESG, E, S and G score changes are often the same for different companies, since Refinitiv usually updates these data annually and often on the same day every year. This means that cumulative average abnormal returns might be influenced to some extent by external factors that cause firms' realized excess returns to deviate from the expected excess returns at certain points in time, and that this deviation might not be directly linked to a specific event. It is especially relevant to analyse whether there are other factors which might be affecting stock prices immediately after the 1st of January each year, since a substantial number of event dates in this study coincide on this date.

There is a plenty of academic research dedicated to proving the existence of the "January effect", which was first discovered in 1976 and refers to the tendency for stock prices to increase in January. Various researchers have found evidence of this trend, but it remains uncertain what drives this anomaly. More recent studies, such as AY Gu (2003), state that the existence of this effect has been in decline ever since the stock market crash of 1987 and that this might suggest that markets are becoming more efficiency. Conversely, Moller and Zilca (2008) defend that this anomaly remains as strong today as it has been in the past. They find that there is a decline in the January effect towards the end of the month, with higher abnormal returns registered in the first half and lower abnormal returns in the second one. Thus, in aggregate these two movements offset each other, causing the monthly abnormal returns to appear similar to the rest of the year.

Then, assuming that there is indeed a January effect, this should be taken into consideration when interpreting the main results from the event study. When events take place on January 1st, the event window covers part of the first half of the month when this anomaly is, supposedly, mostly felt. This would suggest that the event study might find larger positive abnormal returns during this month, caused by the general increase in stock prices instead of due to investors' reactions to the changes in ESG ratings. The main results from the event study, however, indicate that both upgrades and downgrades in ESG scores lead to negative cumulative abnormal returns. So, it is possible that the actual abnormal returns are even more strongly negative if we account for the presence of the January effect.

Additionally, I go one step further to try to exclude other possible factors, besides rating changes, that could potentially affect stock market returns around the same period when rating

events occur. In particular, I analyse dividend announcement dates for the S&P 500 companies between 2011 and the end of 2020. Using data from Datastream, I verify that, in total, only 8% of all dividend announcements during this period occurred in January, so it is reasonable to disregard this factor. Moreover, S&P 500 companies are subject to the regulations imposed by the U.S. Securities and Exchange Commission (SEC) and, as such, they are required to file a 10-K report between 60 and 90 days after the end of every fiscal year. This report is available to the public and, consequently, to investors, and its composition is very similar to the annual reports that companies typically present on their websites. Since the end of the fiscal year is not the same for every company, as it is not mandatory that it corresponds to a calendar year, this suggests that the release dates of annual reports are not likely to be concentrated around a specific period and, thus, do not seem to particularly impact stock returns around January 1st.

7. Conclusion

Firms and investors alike are becoming increasingly aware of the importance of matters related to ESG nowadays. Studies have shown that firms can benefit from ESG disclosure and there is a growing number of firms currently providing ESG reports of their practices. Additionally, there is also a growing trend towards ESG investing, with significant literature dedicated to analysing different ESG strategies.

This thesis adds to the existing literature surrounding ESG topics by assessing whether ESG ratings provide new relevant material information to investors that can influence their investment decisions and how this affects stock market returns. In particular, the goal is to determine the impact of ESG rating downgrades and upgrades on firms' stock returns. To do this, I perform an event study to determine if there are abnormal returns surrounding rating events. I distinguish between the environmental, social and governance components because the aggregate ESG rating might hide changes in each of the components.

The main results from the event study suggest that both upgrades and downgrades in ESG, E,S and G scores are responsible for a decrease in firms' stock returns. In terms of the stock market reaction to rating downgrades, the results show statistically significant CAAR in the 5-day period prior to the event day, except for environmental scores, and also negative CAAR in the period after and including the event day, significant for all 4 categories when using the 6-factor model. This is in line with most findings from existing literature.

Regarding rating upgrades, the event study shows negative CAAR across all dimensions, ESG, E, S and G, before and after the event. These findings contrast with much of the research on this topic, which mostly find that either the stock market reacts positively to positive ESG news or that it shows no reaction to this type of news. However, there is at least one piece of research, Krüger (2015), which provides evidence that, in general, investors react negatively to positive news. Furthermore, he suggests that this reaction might stem from the fact that firms suffer from agency problems, especially those with a considerable market cap.

Moreover, when comparing the cumulative average abnormal returns from upgrades and downgrades in ESG, E, S and G scores, it is possible to observe that the magnitude of the stock market's reaction is larger in the case of rating downgrades. While there are authors who find stronger market reactions to positive ESG news than to negative ones, these results from the event study are once more consistent with the findings from Krüger (2015).

The aforementioned results were obtained using an estimation window of 200 trading days. To ensure the robustness of these results, I repeated the event study using a 245-trading-days window and, overall, the outcome is very similar. Additionally, I try to analyse whether abnormal returns are more significant when firms have previously higher or lower ratings. However, the results are mostly not statistically significant, so it is not possible to draw any relevant conclusions in this regard.

Overall, even though this event study provides some statistically significant results, and these appear to be robust, there are nevertheless some constraints to this analysis. Researchers frequently question the reliability of ESG data, which is understandable given a context where there is currently such substantial disagreement in ESG ratings from different rating agencies. Thus, since this thesis deals with ESG data, this is an important issue to take into consideration when contemplating the results. In fact, this work could potentially be enhanced by collecting ESG data from other sources. Given this discrepancy in ratings, if these findings persisted when using data from different providers, then this would strengthen the reliability of the results shown here.

References

Berg, F.; Kölbel, J., Rigobon, R. *Aggregate Confusion: The Divergence of ESG Ratings*. Review of Finance, 26, 1315–1344

Billio, M., Costola, M., Hristova, I., Latino, C., Pelizzon, L., 2021. *Inside the ESG ratings: (dis)agreement and performance*. Corporate Social Responsibility and Environmental Management, 28, 1426–1445.

Brandon, R., Krueger, P., Schmidt, P., 2021. *ESG rating disagreement and stock returns*. Financial Analysts Journal, 77, 104–127.

Capelle-Blancard, G., Petit, A., 2019. *Every Little Helps? ESG News and Stock Market Reaction*. Journal of Business Ethics, 157, 543–565.

Chen, Z., Xie, G., 2022. *ESG disclosure and financial performance: Moderating role of ESG investors.* International Review of Financial Analysis, 83.

French, K.. *Kenneth R. French – Data Library.* Accessed October 2022. Available at <https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html>

Giudice, A., Rigamonti, S., 2020. *Does Audit Improve the Quality of ESG Scores? Evidence from Corporate Misconduct.* Sustainability, 12, 5670.

Glück, M., Hübel, B., Scholz, H., 2022. *ESG Rating Events and Stock Market Reactions*. Working paper, available at SSRN.

Gu, A., 2003. *The declining January effect: evidences from the U.S. equity markets*. The Quarterly Review of Economics and Finance, 43, 395-404.

Jonsdottir, B., Sigurjonsson, T.O., Johannsdottir, L., Wendt, S., 2022. *Barriers to Using ESG Data for Investment Decisions.* Sustainability, 14, 5157.

Krüger, P., 2015. *Corporate goodness and shareholder wealth*. Journal of financial economics, 115, 304-329.

Moller, N., Zinca, S., 2008. *The evolution of the January effect*. Journal of Banking & Finance 32, 447–457.

Refinitiv, 2022. *Environmental, Social And Governance Scores From Refinitiv.* Accessed September 2022. Available at:

<https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf>

Refinitiv. *The Refinitiv Business Classification.* Accessed October 2022. Available at: <https://www.refinitiv.com/content/dam/marketing/en_us/documents/quick-reference-guides/trbc-business-classification-quick-guide.pdf>

Serafeim, G., Yoon, A., 2022. *Which Corporate ESG News Does the Market React To?* Financial Analysts Journal, 78, 59-78.

Shanaev, S., Ghimire, B., 2022. *When ESG meets AAA: The effect of ESG rating changes on stock returns*. Finance Research Letters, 46.

Appendix

CAAR[-5,-1]	ESG scores	E scores	S scores	G scores
1-Factor Model				
Upgrade	-0,0014**	-0,0020***	-0,0025***	-0,0017***
	(-2,53)	(-2,95)	(-4,03)	(-3,13)
Downgrade	-0,0033***	-0,0015	-0,0025***	-0,0030***
	(-4,50)	(-1,58)	(-2,94)	(-4,83)
Unchanged	-0,0024***	-0,0023***	-0,0021***	-0,0021***
	(-5,20)	(-5,01)	(-4,74)	(-3,57)
6-Factor Model				
Upgrade	-0,0017***	-0,0025***	-0,0026***	-0,0021***
	(-3,00)	(-3,78)	(-4,37)	(-3,97)
Downgrade	-0,0034***	-0,0017*	-0,0027***	-0,0029***
	(-4,80)	(-1,78)	(-3,37)	(-4,93)
Unchanged	-0,0025***	-0,0023***	-0,0024***	-0,0022***
	(-5,64)	(-5,23)	(-5,45)	(-3,85)
Market Model				
Upgrade	-0,0013**	-0,0018***	-0,0020***	-0,0015***
	(-2,27)	(-2,61)	(-3,30)	(-2,90)
Downgrade	-0,0032***	-0,0013	-0,0023***	-0,0027***
	(-4,27)	(-1,33)	(-2,73)	(-4,44)
Unchanged	-0,0021***	-0,0021***	-0,0019***	-0,0019***
	(-4,74)	(-4,59)	(-4,35)	(-3,21)

CAAR[0,10]	ESG scores	E scores	S scores	G scores
1-Factor Model				
Upgrade	-0,0038***	-0,0046***	-0,0011	-0,0042***
	(-2,81)	(-3,01)	(-0,76)	(-3,24)
Downgrade	-0,0046*	-0,0080***	-0,0071***	-0,0032*
	(-1,94)	(-2,70)	(-2,98)	(-1,90)
Unchanged	-0,0038***	-0,0027**	-0,0043***	-0,0039***
	(-3,50)	(-2,52)	(-3,89)	(-2,69)
6-Factor Model				
Upgrade	-0,0053***	-0,0069***	-0,0034**	-0,0055***
	(-4,08)	(-4,60)	(-2,43)	(-4,40)
Downgrade	-0,0060***	-0,0080***	-0,0085***	-0,0049***
	(-4,08)	(-2,69)	(-3,52)	(-2,91)
Unchanged	-0,0060***	-0,0045***	-0,0059***	-0,0062***
	(-5,54)	(-4,12)	(-5,43)	(-4,37)
Market Model				
Upgrade	-0,0031**	-0,0035**	-0,0005	-0,0037***
	(-2,34)	(-2,30)	(-0,35)	(-2,89)
Downgrade	-0,0039*	-0,0075**	-0,0063***	-0,0022
	(-1,68)	(-2,56)	(-2,64)	(-1,34)
Unchanged	-0,0030***	-0,0019*	-0,0036***	-0,0030**
	(-2,81)	(-1,80)	(-3,27)	(-2,14)

Table 4.2: CAAR[-5,-1] and CAAR[0,10]) of rating upgrades, downgrades and unchanged ratings with respect to ESG, E, S and G scores ESG, E, S and G scores, using an estimation window of 245 trading days.