Norwegian School of Economics
Bergen, Fall 2020

ESG Score Changes and Stock Price Reactions

An event study of stock market reactions to changes in Thomson Reuters' ESG Score in the Nordic region

Nils Henrik Benske and Ole Morten Kristiansen

Supervisor: Jørgen Haug

Master thesis, Economics and Business Administration

Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

This page was intentionally left blank.

Abstract

This thesis investigates whether ESG Score announcements by Thomson Reuters present new information to investors in the Nordic region. The data set consists of 1278 unique events for 309 unique firms publicly noted in either Sweden, Denmark, Norway or Finland in the time period from 2011 to 2020. ESG Score announcements are viewed relative to the score from the previous year, and segmented into positive, negative, and neutral events based on the magnitude of change. Positive events have a year-on-year change of at least 10.61 percent, while negative events have a year-on-year change of less than -2.77 percent. We apply event study methodology and define a test battery consisting of a parametric and a non-parametric test, yielding a test battery robust to flaws in our data set. We found evidence of positive abnormal stock price reactions in days prior to the announcement of both positive and negative events. Therefore, we find evidence that the ESG Score might bring new information to investors. Furthermore, probit regressions are used to examine whether firm characteristics correlate with stock price reactions. We find evidence that smaller firms have a higher likelihood of positive abnormal stock price reactions in days prior to positive and negative events. Moreover, positive events dated 2017 or later, have a higher likelihood of positive stock price reactions in the days prior to the events. The same effect, meaning a positive relationship between the likelihood of positive abnormal price reactions and events dated 2017 or later, is also present for the day prior to negative events.

Keywords – ESG Score, event study, firm characteristics, probit regression.

Preface

This thesis is written as a part of the MSc degree in Economics and Business Administration at the Norwegian School of Economics (NHH), where both authors are majoring in finance. The process of writing this thesis has been educational both academically, and in terms of learning the extent of writing an academic paper. We are convinced that the learning outcome of writing this thesis has enriched our skill set to meet future challenges after our studies at NHH.

A special thanks to our supervisor, Associate Professor Jørgen Haug, for his valuable insights and constructive criticism throughout the process of writing our thesis.

Norwegian School of Economics

Bergen, December 2020

Nils Henrik Benske

Ole Morten Kristiansen

Contents

List of Figures

List of Tables

1 Introduction

Sustainability has become an increasingly relevant term for investors and firms. The sustainability of a firm is, however, not easily measured. Its interpretation varies, depending on an individual's definition of sustainability as well as the perception of a firm's sustainable initiatives.

Evaluating sustainability in investments is developing from being considered a niche, to an important evaluation point in a majority of investment decisions. Intergovernmental actions, such as the United Nations development goals (United Nations, 2020), are contributing to increase the awareness of sustainability among investors.

For example, mutual funds which evaluate firms on sustainability are predicted to make up almost 60 percent of total mutual fund assets in Europe come 2025 (PricewaterhouseCoopers, 2020). When investors are to include sustainability in their investments decisions, a more common approach to evaluate sustainability is needed. Several agencies have developed metrics to measure the ESG[1] initiatives of a firm, which may be used to help investors evaluate sustainability.

Thus, it becomes interesting to measure the impact each ESG metric has on stock prices. The effect of Thomson Reuters' ESG Score (hereafter ESG Score) is particularly interesting, as Thomson Reuters is one of the largest providers of professional information.

The effect of the ESG Score may vary across geography, and observed stock price reactions could therefore differ depending on the market(s) examined. Nordic countries present interesting markets for examining the impact of the ESG Score for two major reasons. Firstly, implementing sustainability in investments decisions have a long history in the region (Alfred Berg, 2020). Secondly, the Nordic countries have a vision of becoming the most sustainable region in the world by 2030 (Nordic Council, 2020).

This thesis examines the short-term effect of the ESG Score on stock prices in Nordic countries[2]. The thesis is structured as follows: A short explanation of the ESG score and the main methodology applied in this thesis is provided in Section 1.1 and Section 1.2. In Section 2, a literature review of existing research on the topic is presented. Our

[1]ESG stands for Environmental, Social and Governance.

[2]Nordic countries are hereby defined as Sweden, Denmark, Norway, and Finland.

research questions, including hypotheses based on existing literature are presented in Section 3. Furthermore, the main methodology of use is elaborated on in Section 4, before the data of interest is presented in Section 5. Thereafter, empirical findings and analysis are presented in Section 7, and summarised in Section 8.

1.1 ESG Score

According to a report published by Sustainable Insight Capital Management (2016), there are more than 100 ESG rating providers. Agencies such as MSCI, Sustainalytics, Bloomberg, and Thomson Reuters are categorized as the primary rating providers.

With more than 9000 public firms worldwide and 2100 firms in Europe, Thomson Reuters offers one of the largest databases on ESG Scores, and the database has been used in several empirical studies (Ioannou and Serafeim (2010); Cheng et al. (2014); Drempetic et al. (2020)). The ESG Score consists of three equally weighted pillars - environmental, social and governmental – resulting in a metric ranging from zero to one hundred. Environmental refers to the firm's impact on the environment, measured by metrics such as pollution and CO2 emission. The social pillar includes the firm's focus on the working environment and equality. The governance pillar reflects the firm's work on tax strategy, corruption and risk management. Each pillar is measured with more than 450 data points, such as water usage and human rights policies, and all data points are based on public information (Refinitiv, 2020). The ESG Score is published in the Thomson Reuters database once a year.

1.2 Event Study

The event study methodology is used in this thesis. Its main components will be carefully elaborated later on in Section 4, but a short introduction is provided here. In this thesis, event study methodology is is applied to analyse whether an event causes abnormal stock returns. Figure 1.1 illustrates the main components of an event study.

An event happens on a specific date T_0. To capture variation in stock prices related to the event, an event window is defined, which at a minimum contains the date of the event.

Detecting abnormal returns requires a benchmark for normal returns, and the choice of benchmark model varies in event study methodology. Normal returns are, however, usually estimated using estimation window observations. The estimation window is defined as a period prior to the specific event window, and the estimation window may vary in duration.

In economic terms, the event study is conducted to examine if an event causes returns which are more extreme than normal. Normal returns are estimated in the estimation window, before they are compared to actual returns in the event window. Deviations between actual returns and normal returns are defined as abnormal.

To detect whether abnormal returns are significantly different from zero, it is usual to apply one or more statistical tests. The structure of tests vary, which means that different tests have different features.

Figure 1.1: Event Study Timeline

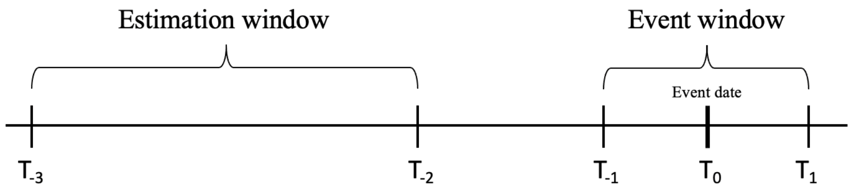


Figure 1.1 displays main components of an event study. T_0 refers to the event date. The event window refers to the time period surrounding the event date, where T_{-1} is the lower limit, and T_1 the upper limit. The estimation window refers to a defined period before the event window, where T_{-2}. is the lower limit, and T_{-3} the upper limit.

1.2.1 Event Definition and Grouping

An event, which falls on an event date, is defined as the announcement of the ESG Score. Whereas Thomson Reuters provides a new score for each announcement, we choose to examine change from the previous score. As the score is usually announced once a year on the same date as the previous year, the events of interest become year-on-year change.

Events are organized in three groups. A distinction is made between a positive event, a neutral event, and a negative event, which all are related to the year-on-year change in ESG Score. The distinction is made in relative terms, meaning a positive event is defined as the upper quartile (inclusive) of total events, whereas a negative event is defined as the lower quartile (inclusive) of total events. A neutral event is naturally defined as the middle ground. Therefore, positive events are defined as an ESG Score year-on-year change of more than 10.61 percent. Negative events are defined as a year-on-year change of less than -2.77 percent. Neutral events are defined as a year-on-year change less than 10.61 percent and more than -2.77 percent.

Grouping is done to examine the whole data set without assuming linearity, as we not expect it to be a linear relationship between the ESG Score and abnormal returns.

2 Literature Review

Findings on the relationship between ESG and the financial performance of firms are divided. The topic has been approached with different methodologies, but resulting in inconsistent results. Some findings show to a positive relationship (Capelle-Blancard and Petit (2017), Brogi and Lagasio (2019)), some negative (Krüger, 2015), while others find no relationship between ESG initiatives and financial performance (Alexander and Buchholz, 1978). This literature review provides an overview of research relevant to our research questions, presented in Section 3.

Historically, research related to ESG initiatives has primary focused on the environmental aspect, such as Bragdon and Marlin (1972). More recent research has also studied social responsibility and governance, meaning the full aspect of the ESG term. For example, Capelle-Blancard and Petit (2017) investigated whether ESG related news affect stock prices.

2.1 Early Stages

Bragdon and Marlin (1972) is an early study on the relationship between environmental performance and firms' financial performance. They investigated whether the profitability of a firm is related to its air and water pollution. A common belief was that firms had to choose between "doing good" and being profitable, where doing good was being environment-friendly. Bragdon and Marlin investigated 17 pulp and paper production firms in the United States from 1965 to 1970, and found a positive relationship between profitability and low pollution – an indication that firms do not need to decide between being profitable and taking care of the environment. An explanation for the relationship was that pollution control reduced labour costs, by reducing health insurance premiums. At the time, pollution caused severe long-term health problems for employees. Additionally, recycling paper contributed to lowering the raw material costs.

A reason for early beliefs of a trade-off between profits and environmental initiatives, was that environmental initiatives were affiliated with higher costs. This claim was made by Vance (1975), but later shelved by Alexander and Buchholz (1978). Alexander and Buchholz investigated whether social responsibility had a positive effect on stock prices, and social responsibility was measured by surveys. The study is based on 40 firms from 1970 to 1974, and results showed no statistical significant relationship between social responsibility and stock returns. They therefore concluded on stock markets to be efficient as described by Fama (1970), which indicates that the value effect of responsible action was already reflected in the stock price. An additional explanation of their findings was that investors were indifferent to firms' approach to social responsibility.

2.2 Announcements

This thesis examines the announcement of ESG Scores. Previous research related to the effect of ESG announcements on stock prices are therefore highly relevant.

Laplante and Lanoie (1994) used event study methodology to analyse how stock prices of 47 Canadian firms reacted to ESG related news. They found no abnormal returns the day after negative environmental incidents were announced, but abnormal returns of -1.2 percent were found the day after a firm announced it would invest in anti-pollution equipment. The latter finding is similar to that of Vance (1975), and indicates that investors believe costs of environmental improvements exceeds expected benefits.

Krüger (2015) used event study methodology to examine the short-term stock price reactions of negative environmental and social news. He found that the news announcements are followed by a short-term decrease in stock price. The study is based on 745 firms from 2001 to 2007. Krüger obtained data from the leading database on ESG information for firms in the United States, which now corresponds to MSCI[3]. Additionally, Krüger shows that a abnormal negative price reaction is present when improvements of corporate social responsibility (CSR) are announced. The latter is similar to the findings of Vance (1975) and Laplante and Lanoie (1994), and indicates that investors find CSR initiatives to be costly, and to exceed expected future benefits. However, governance related news were not accounted for in Kruger's study, due to the doubts of measurement techniques used. The study by does therefore only account for reactions to news related to the environment and social responsibility.

Capelle-Blancard and Petit (2017) used event study methodology to examine short-term stock price reactions to more than 30000 positive and negative ESG related

[3]MSCI is an abbreviation for Morgan Stanley Capital International.

announcements, provided by the Covalence Ethical Quote database. Examples of positive news announcements were winning green awards or launches of environmental friendly products, whereas negative news among others included news about downsizing, toxic release, and bad labour environment. The study is based on 100 public firms included in the Dow Jones Titans index from 2002 to 2010. They found negative abnormal stock price reactions of -0.1 percent after the publication of negative ESG relative news. In terms of the stock price reactions to positive ESG news, the findings of Capelle-Blancard and Petit indicate no relationship.

Brogi and Lagasio (2019) examined the relationship between long-term financial performance of firms and their ESG initiatives. ESG initiatives conducted by each firm were measured by a binary ESG scoring model provided by MSCI[4], whereas financial performance was measured by the firm's return on assets[5]. Brogi and Lagasio analysed more than 17000 observations on 3476 firms in the United States from 2000 to 2016, and their findings indicate a positive relationship between ESG initiatives and ROA, concluding on ESG initiatives to create value for firm stakeholders. In addition, they find differences across sectors. Banks are in particular profiting financially (higher ROA) from focusing on ESG initiatives, compared to other firms. The researchers explain this relationship by United States banks' long focus on lowering their environmental footprint.

Elayan et al. (2014) investigated whether stock prices are affected by ethical performance. The study is based on data from 541 multi-national firms from 2006 to 2009. Ethical performance was measured on quarterly updates from the Covalence Ethical Quote index. Their findings show positive abnormal stock price reactions to ethical upgrades and negative abnormal stock price reactions to ethical downgrades. This indicates that the Covalence Ethical Quote ranking represents new information to investors. Contrary to the findings of Krüger (2015) and Laplante and Lanoie (1994), findings indicate that investors interpret expected benefit of ESG initiatives to exceed expected costs. Another interesting finding is an asymmetric stock price reaction to up- and downgrades. The magnitude of abnormal returns following ethical upgrades were less than the magnitude of abnormal returns following ethical downgrades. This finding indicates that investors seem to punish ethical downgrades more than rewarding upgrades.

[4]The scoring model is based on the MSCI ESG KLD STATS database.

[5]Return on assets (ROA) was calculated as net income divided by total assets.

Elayan et.al. also investigated whether firm characteristics are associated with positive or negative stock price reactions. The results show that financially well-performing firms are associated with positive price reactions following an Ethical Score upgrade. Financial performance was measured by a metric developed by the researchers[6].

As stated initially in the literature review, a clear consensus on how ESG related news affect stock prices does not exist. Event study methodology has been widely used, for example by Laplante and Lanoie (1994), Elayan et al. (2014), and Krüger (2015). Another popular approach not mentioned above is index inclusion and exclusion (Cheung (2011), Oberndorfer et al. (2011)). Similar to ESG announcements, mentioned research on index inclusion and exclusion do not provide any clear consensus in terms stock price reactions to ESG news.

[6]The metric consists of 13 different financial metrics, such as return on assets, return on equity, and earnings per share. See Elayan et al. (2014) page 389 for a more detailed description.

3 Research Questions and Hypotheses

In this section, we define research questions and hypotheses. Primarily, event study methodology is applied to investigate stock price reactions around ESG Score announcements. In addition, we investigate the correlation of firm characteristics with stock price reactions around names announcements.

3.1 Research Questions

Evaluation is performed by examining two research questions, which aim to examine whether the ESG Scores provide new information to investors, and the correlation of firm characteristics with stock price reactions. Plantinga et al. (2015) find that due to time and resource restrictions, an investor will not be able to process all the available information on ESG initiatives, and will therefore rely on ESG scores provided by ESG rating agencies. As a result, an argument can be made that ESG Scores are an important metric for investment decisions.

Our main research question is defined by:

Research Question 1. Do announcements of the ESG Score contain new information for investors?

In addition to the main research question, we will examine characteristics of firms which experiences positive and negative stock price reactions around events. More specifically, we seek to examine whether there are firm characteristics which increases the likelihood of a firm experiencing a positive or negative stock price reaction. The second research question is defined as:

Research Question 2. Do price reactions to ESG announcements correlate with firm characteristics?

3.2 Hypotheses

In order to investigate our research questions, we define four hypotheses and what we expect to find.

Hypothesis 1. Announcements of positive events are associated with abnormal stock price reactions.

We hypothesise that the ESG Score announcements represents new information to investors. Therefore, we expect abnormal price reactions as investors react to the announcements.

As discussed in the literature review in Section 2, existing literature is divided in terms of the direction of stock price reactions to ESG initiatives. Hence, we do not hypothesise the direction of abnormal stock price reactions to positive events, only that there will be abnormal stock price reactions.

Hypothesis 2. Announcements of negative events are associated with abnormal stock price reactions.

On the same note as Hypothesis 1, we hypothesise that there are abnormal price reactions around negative events. Once again, the hypothesis is that there will be abnormal price reactions, and no specific direction is predicted. This notion is based on existing literature, already presented in Section 2.

Hypothesis 3. Announcements of neutral events are not associated with abnormal stock price reactions.

We hypothesise that there are abnormal price reactions around positive and negative events, and that there are no abnormal price reactions around neutral events.

Therefore, we primarily aim to examine positive and negative events. Neutral events are best viewed as a control group, where no abnormal price reactions are expected.

Hypothesis 4. Certain firm characteristics are correlated with positive and/or negative stock price reactions.

We hypothesise that there are certain firm characteristics which are correlated with the likelihood of positive abnormal price reactions, given a positive or negative event. It is expected as we believe firm characteristics affects how firms are positioned to capitalize on ESG initiatives, or lack of ESG initiatives.

4 Theoretical Framework

Initially, early work within event study methodology is presented in this section. Furthermore, most main components of an event study, which were already introduced in Section 1.2, are elaborated on. The components are applied to provide a framework for answering Research Question 1 and hypotheses as defined in Section 3.

Event studies have been a popular tool to examine events for publicly listed firms for decades. The intuition behind event studies may be seen in relation to the efficient market hypothesis (Fama, 1970), which in its strong form states that stock prices reflect all available information. If the efficient market hypothesis holds, then stock prices should only react to new information. If markets are semi-strong, stock prices should reflect new information quickly, which usually is the basis for event studies.

Ball and Brown (1968) examined the effect of accounting earnings announcements on stock prices. They based their analysis on earnings data in the United States from 1946 to 1966, as announced by the Wall Street Journal. As they found no abnormal price reactions, they concluded on earnings information already being incorporated in stock prices.

4.1 Estimation Window

An important component in event study methodology is the estimation window[7]. As introduced, its purpose is to estimate normal returns for observations in the event window. The estimation window length is chosen balancing increased precision (longer window) with including as little unrelated movements as possible (shorter window). A popular choice in literature is an estimation window ranging from 250 days before the event to 30 days before the event, but there is no set length for estimation windows (Aktas et al., 2007).

Setting an appropriate length for the estimation window is important to obtain valid results. Poor choice of estimation window could cause biased normal returns. However, the estimation window has, somewhat surprisingly, historically attracted less interest in

[7]Illustrated in Section 1.2.

event study methodology (Aktas et al., 2007).

The estimation window should not include variation tied to the event window, which is the reason researchers set the estimation window to end before the event window (MacKinlay, 1997). Also, Ball and Brown (1968) and Fama et al. (1969) states that the estimation window of an event study should not include the event window, as it would cause biased normal returns. To understand this bias, consider again the purpose of an estimation window. Its purpose is to estimate normal returns, which in turn are used to detect whether an event induces abnormal returns. Including event-specific variation in the estimation window would violate the separation of normal and abnormal returns, and thus cause biased normal returns.

Another bias which could be present in the estimation window is in the form of including unrelated stock movements. The intuition is that including observations which are not relevant would interact with standard errors of normal returns, for example by overestimating standard errors if unrelated movements included deviates heavily from the mean of estimation window returns. Overestimating standard errors would cause a type 2 error when testing for abnormal returns, meaning a false negative. A false negative occurs when the null hypothesis is kept when it, in reality, should be rejected. As standard errors usually are in the denominator for in test statistics for abnormal returns, overstating them understates the test statistic, leading to type 2 errors.

A solution to this problem is to manually exclude unrelated movements from the estimation window, which proves laborious for medium to large samples. Another option is to test for abnormal returns using a test which aims to exclude unrelated movements in the estimation window, such as the test of Aktas et al. (2007). A third option is to use Winsorizing[8], a technique where the magnitude of extreme outliers are reduced.

Although sounding promising, trying to exclude unrelated stock movements might actually hurt more than it improves in terms of event study specification. The reason is that it is hard to determine what is related stock price movements in the estimation window and what is not. Attempting to single out unrelated stock price movements could therefore cause unexpected biases, so neither of the three mentioned options for eliminating unrelated stock price movements in the estimation window are applied in this thesis.

[8]Named after Charles P. Winsor.

We set the estimation window length to [-200, -10][9], meaning it starts (inclusive) 200 days before the ESG Score announcement and ends (inclusive) 10 days before the ESG Score announcement.

4.2 Event Window

Defining the event window is another important step to conduct an event study. MacKinlay (1997) states that the event window often includes at least the event day and the following day. Event windows could be expanded to include days prior to the event day if information leakage is suspected (MacKinlay, 1997), and more days after the event day if it is necessary to capture post-event day returns related to the event. The intuition behind setting the event window length is to include all event specific information while excluding information not related to the event. In practice, that is hard to do, as there is no such thing as "one effect present" in the stock market. The stock market is complex and continually affected by many effects, which makes an assumption that the event is the only factor which affects stock prices at a specific time point weak. The solution is a trade-off, where extending the event window length leads to capturing more information, both event specific and non-related information.

Setting the event window length, thus includes a question of how fast the market reacts to new information. If the market is semi-efficient according to Fama (1970), this period should be short for there to be no arbitrage opportunities. However, the duration of market imbalance following the news, which is defined by Krivin et al. (2003) as the duration where future stock movements are affected by past news, has been studied ever since Fama (1970). Research on the topic includes Hillmer and Yu (1979) and Chordia et al. (2005), who finds that the market imbalance period lasts for a matter of hours or minutes, respectively, thus making a case for the use of a single-day event window. Krivin et al. (2003) states that the imbalance period is useful to set a minimum duration of the event window, but it does not necessarily define the proper length.

Lev (1989) examines 19 event studies conducted and published in four major accounting

[9]Ideally, a thorough analysis should be conducted for stock markets in Norway, Sweden, Finland, and Denmark to determine the optimal estimation window duration for events over our time span, but it is laborious and assumed to be of minimal interest for the reader. Therefore, the estimation window is set somewhat arbitrary, but within the usual limits in event study literature.

journals in the 1980s where the vast majority use fixed event windows (same for all events in the study), with lengths ranging from two days to a year. A popular choice of event window length in more recent literature has been a three-day event window (Campbell et al. (2010); Capelle-Blancard and Petit (2017)), or multiple days before and after the event (Krüger (2015); Elayan et al. (2014)).

The choice of event window length determines a distinction in event study literature. The distinction between short-term event studies and long-term event studies is not set in stone, but is usually done around the one year mark ((Kothari and Warner, 2007)). Hence, a long-term event study examines stock prices for a duration of at least a year around the event, while a short-term event study examines stock prices over a shorter time span, often just a couple of days. While Fama et al. (1969) is a long-term event study, more recent literature generally leans on short-term event studies. Short term event studies are considered to provide clear evidence when specified correctly (Fama, 1991), while the interpretation of long-term event studies are more problematic (Kothari and Warner, 2007). Fama (1998) finds that long run anomalies cannot be replicated using a variety of statistical approaches, and they are therefore found by chance. Kothari and Warner (1997) find that inferences in long-horizon studies are not reliable, while Lyon et al. (1999) finds that long-horizon studies are sensitive to non-random sampling. Literature generally find results of short-horizon tests more dependable than those of long-horizon tests (Brown and Warner (1980); Kothari and Warner (2007))

If investors, in line with the findings of Plantinga et al. (2015), use the ESG Score to interpret complex information, they would be interested in acquiring the information before other investors. Hence, abnormal returns due to leaked information might be present in our study.

ESG Score announcements differ from other typical firm announcements examined in event studies such as earnings announcements. ESG Scores do not represent insider information and are not subject to strict regulations in terms of notifying all stakeholders at the same time. Therefore, it is possible that ESG Scores are treated less carefully, and information is more likely to be leaked.

For robustness purposes, we define ten different event windows in our study. Three event windows are defined to capture abnormal returns due to leaked information, namely event windows [-5,0], [-3,0], and [-1,0]. Furthermore, four event windows are defined to capture effects after the ESG Score announcement. The four event windows are [0,1], [0,3], [0,5], and [0,10]. Three additional event windows are defined to capture effects on both sides of the ESG Score announcement, namely event windows [-5,5], [-3,3], and [-1,1]. Event window duration ranges from 2 days to 11 days, and days are defined as trading days, meaning days where the respective stock exchange is open. The event day is either the day of the ESG Score announcement or the first trading day after.

Note that all event windows are kept short to avoid the problems of long-horizon event studies, but all event windows are over multiple days to capture effects of the announcement around the event day.

4.3 Benchmark Return Model

The estimation of normal returns, illustrated in Section 1.2, requires a benchmark return model. Quite a few different models have been applied in event study literature. Statistical methods include the constant mean model and the single-factor market model, while economic models include the Capital Asset Pricing Model (CAPM), the Fama-French factor models and the Arbitrage Pricing Theory (APT) model.

The single-factor market model is chosen to estimate normal returns in this thesis. The market model is chosen as other, often more complicated, models do not outperform the market model (Brown and Warner (1980); Binder (1998)). A clear benefit of the market model compared to other statistical models such as the constant mean model, is that the market model incorporates each stock's sensitivity to the market return. It follows that the market model is more robust compared to the constant mean model in a case where cross-correlation[10] is present. Another feature of the market model, compared to economic models, is that it relies on statistical assumptions, and is not restricted by assumptions related to investor behaviour.

The market model is defined as (MacKinlay, 1997):

$$R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{it}, \tag{4.1}$$

[10]Cross-correlation is elaborated on in Section 6.1.

$$E(\epsilon_{it}) = 0, \qquad Var(\epsilon_{it}) = \sigma_\epsilon^2, \tag{4.2}$$

where R_{it} is the calculated normal return for security i at time t and R_{mt} is the market return at time t. α_i and β_i are regression parameters and are estimated using Ordinary Least Squares (OLS) regression. β_i is each stock's sensitivity to the market index. The regression parameters are estimated over the estimation window.

The residual ϵ_{it} is an estimator of abnormal returns, and the regression method seeks to isolate firm specific information in the error term (Fama et al., 1969). Calculated abnormal returns, as illustrated in Equation 4.2, will be jointly normally distributed with a zero conditional mean and conditional variance (MacKinlay, 1997).

Abnormal returns are defined as the difference between actual returns for security i and the estimated normal returns for security i. Abnormal returns are calculated for each day in the event window, and given by:

$$AR_{it} = R_{it} - R_{mt}, \tag{4.3}$$

where R_{it} is the actual return for security i at time t, and R_{mt} is the estimated normal return for security m at time t. Combining the linear equation for normal return and the equation for abnormal return gives:

$$AR_{it} = R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{mt}. \tag{4.4}$$

In order to calculate reliable normal returns, regression parameters need to be unbiased. Unbiased estimators are achieved when the general assumptions for OLS hold. MacKinlay (1997) argues that it is reasonable to believe that the assumptions hold and rarely lead to any problems. Assumptions are formally checked for the Nordic markets, and they seem reasonable also in this case. Results from a thorough approach is available to the reader in Appendix A1.

To test whether ESG Score changes have an overall effect on stock price reactions, abnormal returns are aggregated across time and securities. Aggregation has two major advantages. One is that aggregation across securities is useful as it is easier to separate a general score effect from firm specific effects. Furthermore, aggregation across securities offsets potential problems related to market misinterpretation of a few individual firms, as the average reaction to an event is measured for a diverse group of firms. Advantage two is that event dates spread across time reduces correlation between events, as well as reduces the problem of systematic errors.

Firstly, abnormal returns are aggregated per event window for each event by defining cumulative abnormal returns by:

$$CAR_{iL} = \sum_{i=1}^{L} AR_{it}, \tag{4.5}$$

where AR_{it} are abnormal returns for security i at time t in the event window and L the upper limit for t defined by the size of the event window.

The last step before testing for statistical significance is to account for the number of events by calculating the average cumulative abnormal return by the following formula:

$$ACAR_{NL} = \frac{\sum_{i=1}^{N} CAR_i}{N}, \tag{4.6}$$

where N is the number of examined events.

4.4 Estimating Market Return

While early studies such as (Fama et al., 1969) focused on a single market, more recent event studies also examine several markets. In this thesis, stock price reactions to ESG Score announcements are examined for Sweden, Denmark, Norway, and Finland, thus making this a multi-country event study. In such event studies, there are likely to be different individual effects for each country. Examples of individual effects that can vary between countries include interest rates, regulations, and currency effects.

In general, the market model specifies that the return of a security depends on its beta[11] multiplied by the market return. A proxy is applied to estimate market return, which is usually defined as the market portfolio. The market portfolio is defined as a bundle of assets which provides the same expected return as the market. A potential proxy is a value-weighted index, which weights each asset equal to its size in the market. Value-

[11]Defined in Equation 4.1.

weighted indices are frequently used in event study literature, and applied by for example Brown and Warner (1985) and Campbell et al. (2010) to represent the market portfolio and thus market return.

However, the market portfolio is based on the notion that it replicates market return. A value-weighted index for the whole market might therefore be imprecise if ESG Scores are not applied to all firms in the market. This depends on the definition of the market. For example, if the researcher defines the Norwegian market as shares noted on Oslo Børs, a value-weighted index for Oslo Børs would present an excellent option for the market portfolio. On the other hand, if Euronext Expand and Euronext Growth are also included in the definition of the market, an index based solely on Oslo Børs would be less precise.

The findings of Campbell et al. (2010) support this argument. They find that a single-market model with national market index works well as a proxy for the market return in multi-country event studies. For this reason, national indices are applied in our benchmark model, the market model presented in Equation 4.1.

Another potential proxy for the market portfolio would be to construct national value-weighted indices for firms assigned an ESG Score, but it falls outside of the scope of this thesis. Instead, we aim to identify the best proxy - which makes value-weighted indices generally attractive. However, two characteristics in the data set stands out[12]: (i) ESG Scores are mainly assigned to large-cap firms, and (ii) some countries have a high representation of a specific industry.

To identify the best proxy for market return for each country, i.e. the market portfolio, several indices based on market capitalisation and industry are tested. As these results are not the main focus of this thesis, results are presented in Table A1.2 in Appendix A1. Some indices present results similar to that of the value-weighted index for each country, but no indices stand out in that regard. In line with Brown and Warner (1985) and Campbell et al. (2010), national value-weighted indices are therefore applied as proxies for the market portfolio.

[12]Firm characteristics are described in Section 5.2. Here, we are content with two main features.

5 Data

This section presents the data used in this thesis. Firstly, the collection and cleaning process are briefly explained in Section 5.1, before descriptive statistics are presented in Section 5.2.

5.1 Data Collection

ESG Scores for public firms in the Nordic from 2010 to 2020 were retrieved from the Thomson Reuters Refinitiv database. Firms were required to have a minimum of two subsequent ESG Score observations in this time span, as ESG Score changes relative to the previous year are of interest in this thesis. The data set for the event study thus consists of 1278 unique events and 309 unique firms. Additionally, data on firm characteristics were retrieved from the same database.

The data set used to examine firm characteristics consists of 612 events and 219 unique firms. Note that it is based on the same data as the event study, but the correlation between firm characteristics and stock price reactions are only examined for positive and negative events. Hence, there are fewer observations. Moreover, a few events were excluded as data on firm characteristics for these firms was not available.

Daily adjusted stock prices[13] for firms were also retrieved. Stock prices for firms listed after 2010 were retrieved from their first trading day. Although early studies, such as Fama et al. (1969), used monthly return data, daily data is used by default in modern event studies. Daily data increases precision compared to monthly data, and Brown and Warner (1985) found that daily data generally do not hurt the specification of event studies.

Index data for calculation of market returns for firms listed in Sweden, Denmark, and Finland was retrieved directly from Nasdaq. For firms listed in Norway, index data was retrieved from Oslo Børs.

To calculate stock and market returns, both simple returns and the natural logarithm of returns may be used. Wooldridge (2013) finds the difference between simple and

[13]Daily adjusted stock prices are stock prices adjusted for dividends, share splits, and other incidents.

logarithmic returns to be very small when results are hovering around zero, which is generally the case for our data set.

The difference of simple and logarithmic returns are examined, and differences are, as Wooldridge (2013) predicts, minor. Figure A2.1 in Appendix A2 substantiates this. Simple returns are used in this thesis.

Simple returns are calculated by the following formula:

$$Simple\ return = \frac{Today's\ price - Yesterday's\ price}{Yesterday's\ price}. \tag{5.1}$$

5.2 Descriptive Statistics

Events are segmented by market capitalisation and industry, based on the firm which has the event. Events are segmented by market capitalisation into small-cap, mid-cap and large-cap events. The segmentation is based on guidelines provided by Nasdaq Nordic (2019) and Oslo Børs (2020). Nasdaq guidelines are applied to the Swedish, Finnish, and Danish markets, while Oslo Børs guidelines are applied to the Norwegian market. Guidelines are presented in Figure A2.1 in Appendix A2.

Descriptive statistics for total events are summarized in Table 5.1. Industry segmentation follows Thomson Reuters standards, which are presented in Table A2.4 in Appendix A2.

Table 5.1: Descriptive Firm Statistics by Observed ESG Score Changes.

	Norway	Sweden	Finland	Denmark	Total
Total events	239	560	225	254	1278
Positive	68	137	50	61	316
Negative	61	142	52	67	322
Neutral	110	281	123	126	640
Market capitalization					
Small-Cap	21	14	1	9	45
Mid-Cap	86	71	24	51	232
Large-Cap	132	475	200	194	1001
Industry segmentation					
Financials	35	85	9	46	175
Energy	96	9	9	9	123
Industrials	24	164	78	82	348
Healthcare	1	36	17	60	114
Materials	21	51	45	18	135
Utilities	1	0	9	4	14
Technology	30	70	32	4	136
CC	25	28	15	12	80
CNS	4	68	10	19	101
Real Estate	2	49	1	0	52

Table 5.1 summarises descriptive statistics of observed ESG rating changes (events) from 2011 to 2020. Firms are traded at the Norwegian, Swedish, Danish, and Finish stock exchange. Events are categorized in three groups: Positive, negative, and neutral. Positive events are identified with an ESG score year-on-year change of more than 10.61%. Negative events are identified with y/y change of -2.77%. Neutral events are identified as y/y change lower than 10.61% and higher than -2.77%. Market capitalisation segmentation is done by applying stock exchange standards (See Table A2.1 in Appendix A2 for a more detailed description). Industry segmentation follows Thomson Reuters standards (See Table A2.4 in Appendix A2 for a more detailed description). CS = Consumer Cyclicals, CNS = Consumer Non-Cyclicals.

Table 5.1 summarises descriptive statistics for positive, negative, and neutral events. The table presents a total of 1278 events, where 316 are categorized as positive, 322 negative, and 640 as neutral. A few events from the original grouping[14] are removed as their stock returns were not available.

Sweden has the most events in the data set, which is sensible given that Nasdaq Stockholm is the largest stock exchange in the Nordic. Sweden has a total of 560 ESG score announcements with 137 positive, 142 negative, and 281 neutral events. In other words, the representation is almost symmetrical across event groups. Norway, Finland, and

[14]See grouping definition in Section 1.2.

Denmark are about equally represented for total events as well as for each event group. ESG Scores in Norway appears to be a bit more volatile than its Danish and Finnish counterparts, as fewer events are categorized as neutral.

Furthermore, large-cap firms are more represented than mid-cap and small-cap firms in the data set. It applies to all four markets, but Finland stands out with 200 of 225 events. In terms of industry description, Industrials is the most represented industry with a total of 348 events. Additionally, Financials and Materials also stand out in terms of representation with 175 and 135 events, respectively. Sweden is clearly dominated by the Industrials industry (164 out of 560), while Norway has a larger share of events for firms in the Energy industry (96 out of 239) compared to the other countries. Finland and Denmark are more evenly-distributed across industries.

The overall impression is an overweight of events from Sweden, and an overweight of large-cap firms.

Table 5.2 is provided to illustrate how events are spread out across time. Note that the base is quite stable for all countries except Sweden in the time period 2011 from 2020, which means that number of events are somewhat stable until 2020. The year 2019 has the most events with a total of 253, which is an increase of 141 events compared to the observations in 2011. 2019 also has the steepest increase from the previous year in number of observations (up to 250 from 159 events in 2018).

Table 5.2: Number of Observed Events by Year.

Year	Norway	Sweden	Finland	Denmark	Total
2011	21	44	22	25	112
2012	21	46	23	25	115
2013	21	48	23	25	117
2014	22	49	24	25	120
2015	22	52	24	26	124
2016	23	61	24	27	135
2017	23	68	24	28	143
2018	28	74	26	31	159
2019	58	116	35	41	250
2020	0	2	0	1	3

Table 5.2 summarises number of total events by year for each country. An event is defined as the announcement of ESG Score.

6 Defining Tests for Abnormal Returns

Most main components of an event study were elaborated on in Section 4. This section will discuss the remaining main component, more specifically the statistical tests used to examine whether abnormal returns are present or not. This section also provides insight to potential problems in the data set, which is used to define a test battery.

6.1 Dependency Problems

In the data set used in this event study, ESG Scores are announced for many firms on the same day. It causes what in event study methodology is called event date clustering. As a result, cross-sectional correlation is present, which is a potential problem (Kolari and Pynnönen, 2010). Salinger (1992) finds that ignoring cross-sectional correlation can lead to seriously underestimated standard errors, while Kolari and Pynnönen find that even low cross-correlation hurts test specification in terms of increasing type 1 errors.

A type 1 error in statistics is a false positive, meaning the null hypothesis is rejected when there is not enough evidence to reject it. To illustrate ways cross-correlation could lead to type 1 errors, consider the situation of event date clustering. In the days surrounding the clustered event day, firms are affected by many co-founding effects. An example could be that the national government just issued a stimulus package for all firms. Thus, if all firms benefit from the stimulus package, one would expect their stock prices to correlate upwards. If this effect is present in the estimation window, normal return calculations would be biased. Similarly, if this effect is present in the event window, returns could appear abnormal due to the stimulus package. Usually, such effects are inevitable in event studies, but event date clustering makes them problematic. Intuition is that in the presence of event date clustering, the mentioned effect would affect many firms, contrary to only one (or a few) firms if event date clustering is not present. Hence, test results are likely to be biased in the presence of event data clustering, if it is not accounted for.

The main insight for dealing with clustered events, is that returns across securities tend to correlate. The intuition applies for both the event window and the estimation window. If test statistics do not account for this phenomena, standard errors are underestimated by assuming zero cross-sectional correlation for returns. As standard errors usually are found in the denominator of test statistics, the understatement leads to a lesser-than-actual value in the denominator. As a consequence, the test statistic is overestimated, leading to false rejections of the null hypothesis of no abnormal returns - type 1 errors.

The problem of classic statistical tests for abnormal returns is that they assume independence in the error term of observations, meaning no cross-correlation. Kolari and Pynnönen (2010) survey all event studies from 1980 to 2007 in four major financial journals, and accounts for their way of coping with cross-correlation. As much as 55 (of 76) used the portfolio method proposed by Jaffe (1974), which aggregates a portfolio of equally weighted securities, to cope with cross-correlation issues. Another solution is to apply a test statistic which is robust to cross-correlation, which is chosen for this thesis.

Event date clustering in the data set is illustrated in Table 6.1. For example, a major portion of events occur on the 31st of December, meaning the first trading day after the event is the first trading day of a new year[15].

Table 6.1: Clustering of ESG Scores.

Dates with 1 event	Dates with 2-5 events	Dates with $>$ 100 events
45	15	9

Table 6.1 illustrates the level of clustering for event dates in the time period 2011 to 2020. 45 dates contains only one event, 15 dates contains 2-5 events, and 9 dates contains more than 100 events.

Country-clustering is another type of clustering effect which could lead to biased results when drawing conclusions for the whole population. As seen from the descriptive statistics presented in Table 5.1, firms traded on the Swedish stock exchange stand out with 560 total events. This indicates a share of almost 50 percent of total observations. One could therefore argue that observations for Sweden are over-represented in our sample. However, Nordic countries are known for their similar governance structure. For example, Sinani et al. (2008) finds that Denmark, Norway, and Sweden are homogeneous in terms of political stability, control of corruption, and accountability. It is therefore reasonable to believe that investors invested in the Nordic market respond similar to ESG related publications. Hence, we argue that the relatively high share of Swedish observations is

[15] Unusual (relatively to the rest of the year) trading patterns around the turn of the year is well-known in finance. For example, related to the phenomena that investors sell stocks at the end of the year for tax (shield) purposes.

not a problem when drawing conclusions for the whole population.

6.2 Event Induced Variance

A reason why the common portfolio method by Jaffe (1974) is not applied in this thesis, is that it is inefficient and not robust in the case of event induced variance (Kolari and Pynnönen, 2010).

Event induced variance is present when the variance of returns increases around the event date. This increase occurs when firms react differently to the new information presented in the event (Boehmer et al., 1991), and event induced variance is especially applicable when operating with daily returns (Brown and Warner, 1985). Ignoring the effect of induced variance typically leads to type 1 errors. Event induced variance becomes a problem when applying statistical tests to detect abnormal returns, as a bias occurs in the comparison of normal and actual returns. Induced variance in the event window causes estimation window observations, which is used to calculate normal returns, to have understated variance relatively to event window returns. Therefore, event window observations appear more extreme than they are in reality, and type 1 errors occur.

ESG Score announcements, and therefore yearly change of ESG Score, may differ substantially in terms of magnitude of year-on-year change. However, even for an identical year-on-year change, there is reason to believe that firms are affected differently. This follows the fact that investor preferences are differentiated, both for individual stocks and across industries. One might for example suspect that a firm in a sector focused on sustainability will have a different stock market reaction to a notable upgrade in a firm's ESG Score, compared to a notable upgrade for a firm in another sector. Furthermore, investors might be more reluctant to invest in a low-scoring FMCG[16] firm than a low-scoring oil firm, if investors are more concerned of the sustainability of the latter than the former.

Event-induced stock price movements are the whole basis of an event study, as the methodology is based on how large abnormal returns are needed to detect statistically significant abnormal returns. Harrington and Shrider (2007) found that all events induce variance, making it a potential problem which is accounted for in this thesis by using

[16]Fast-Moving Consumer Goods

robust tests.

Harrington and Schrider found that increased variance in the event window causes heteroscedasticity in cross-sectional regressions, and thus causing biased estimators for ordinary least squares (OLS) estimation. A fairly standard solution applied by researchers to cope with heteroscedasticity is to use robust standard errors[17].

In this thesis, robust standard errors are not applied. Instead, tests which are robust for event induced variance are used, meaning the test developers have applied robustness methods themselves. Using tests which are robust for event induced variance is consistent with the recommendations of Harrington and Shrider (2007). Also, Kolari and Pynnönen (2010) found that the choice of test statistic is important to avoid type 1 errors as caused by event induced variance.

6.3 Skewness and Kurtosis

Skewness and kurtosis are phenomenons which could affect the proper choice of statistical significance tests, as it affects the data distribution. If skewness and kurtosis are present, a test which does not assume a specific distribution would be more appealing. Skewness refers to lack of symmetry in the data, while kurtosis is present if data are heavy-tailed or heavy-peaked, meaning there are either many extreme observations or an extreme amount of similar observations[18]. Keep in mind the structure of an event study, defined in Section 1.2. Estimation window observations are used to construct normal returns, before event window observations are tested in relation to normal returns[19]. Returns in the event windows do not appear to suffer from skewness, but kurtosis might be present as events are heavily peaked. It is illustrated in Figure A2.1 in Appendix A2.

[17]Robust standard errors could be applied by several methods, for example White's standard errors or Hubard-White standard errors.

[18]Heavy-tailed refers to the two ends of the data distribution, and heavy-peaked refers to the highest point of the data distribution. For example, the Laplace distribution is a heavy-peaked distribution, while the Cauchy distribution is a heavy-tailed distribution.

[19]Assumptions for normal return estimation are discussed in Section 4.3. Subsequently, this paragraph refers to skewness and kurtosis in the event window.

6.4 Test Battery

In this section, we define a test battery for our empirical analysis. Moreover, this section concerns aspects of statistical significance tests in event studies. In general, cross-correlation and event induced variance needs to be accounted for, as discussed in Section 6.1 and 6.2.

As mentioned in Section 1.2, a statistical significance test is used to detect whether abnormal returns are different from zero. There are two main groups of tests, parametric and non-parametric tests. Parametric tests requires the data to follow a particular distribution, whereas a non-parametric test may be viewed as distribution-free tests. Kolari and Pynnonen (2011) states that non-parametric tests are preferable to parametric tests in event study methodology, as stock returns do not follow a specific distribution. The drawback is that a non-parametric test might be less powerful than a parametric test when the data distribution requirement is met, meaning type 2 errors probability is increased.

Following our check for normality for positive, negative, and neutral events for the event window [-3,3], it is sensible to assume that the data follows the Gaussian normal distribution. A parametric test may therefore make sense to use. However, we are testing for several event windows, and also have potential problems in the form of event induced variance, cross-correlation and kurtosis.

As a consequence, we define a test battery consisting of a parametric and a non-parametric test. The battery is modelled by two tests which both are sensible choices given our data set, to ensure robustness in our analysis.

Applying several tests is consistent with the findings of (Campbell et al., 2010), who test the performance of event study methodology in 54 different countries from 1986 to 2006. If test results differ, interpretation should be conducted with caution.

Popular parametric tests include the original BMP test (Boehmer et al., 1991) and Patell test (Patell, 1976). Kolari and Pynnönen (2010) found the original BMP test creating far less type 1 errors compared to the commonly used Patell test when induced variance was present.

Kolari and Pynnönen presented an adjustment of the BMP test, which performs well in the presence of cross-correlation and event-induced variance. The adjusted BMP test (Kolari and Pynnönen, 2010) will therefore be applied in this thesis, and its features are elaborated on in Section 7.1.

A non-parametric test which has been proven effective by several studies is the rank test, proposed by Corrado (1989) and adjusted by Corrado and Zivney (1992). Several studies found that it outperforms parametric tests in terms of specification, for example Campbell and Wesley (1993), who recommended using the rank test for NASDAQ (USA) samples. Moreover, Maynes and Rumsey (1993) found, using Toronto SE data, that the rank test works well for all trading frequencies. Aktas et al. (2007) recommends the rank test as a good choice if skewness and kurtosis are present.

The rank test is intriguing due to its empirically proven performance, but it were originally designed to test for a single-day. It may be aggregated, but still loses power for more days in the event window. We therefore apply an adjustment of the rank test, called the GRANK T-test (Kolari and Pynnonen, 2011). It is robust for longer event windows and for event induced variance, as well as cross-correlation and serial-correlation of returns. It is included as the second part of the test battery.

Hence, the test battery consists of two tests: The adjusted BMP test and the GRANK T-test. Results from the test battery will provide a solid basis to evaluate short term effects of the ESG Score announcements.

7 Empirical Analysis

This section contains our empirical analysis of the ESG Score, and is structured as follows: The adjusted BMP test is presented in section 7.1, before results of the test and interpretation of the results are provided. Section 7.2 includes the definition of GRANK T-test and the results using this test. Thirdly, Section 7.3 includes discussion of results from the test battery as a whole, and is used to answer Hypothesis 1, 2, and 3.

Finally, the correlation between firm characteristics and the direction of stock price reactions are examined using probit regression. This analysis, presented in Section 7.4, is used to answer Hypothesis 4.

7.1 The Adjusted BMP Test

The adjusted BMP test has its origin from the original BMP test (Boehmer et al., 1991), which accounts for event induced variance. In order to reduce the impact of highly volatile returns when averaging abnormal returns, the original BMP test uses standardized abnormal returns. Standardized abnormal returns weights highly volatile observations less than other observations, by taking the observed abnormal returns divided by their inverse standard deviation. Standardized abnormal returns is computed by:

$$A_{it} = \frac{AR_{it}}{s_i\sqrt{1+d_t}}, \tag{7.1}$$

where A_{it} is the standardized abnormal return for security i at time t and AR_{it} is the actual abnormal return for security i at time t. The regression residual standard deviation for security i and d is given by s_i, containing the forecasted correction term estimated by the regression parameters in the estimation window for the event window.

The original BMP test statistic (Boehmer et al., 1991) is given by:

$$t_B = \frac{\overline{A}\sqrt{n}}{s}, \tag{7.2}$$

where $\overline{A}$ is average standardized abnormal returns from Equation 7.1 across events in the event window, n the sample size, and s the cross-sectional standard deviation of standardized abnormal returns, which is the squared root of the variance inside the event window, given by:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(A_i - \overline{A})^2. \tag{7.3}$$

Although the original BMP test contributed to the existing literature by its simple implementation, the test still assumes abnormal returns to be cross-sectionally uncorrelated (Boehmer et al., 1991). As discussed in Section 6.1, this assumption causes type 1 errors when event dates are clustered.

Kolari and Pynnönen (2010) claim to have resolved the weakness of the original BMP test. Their adjusted BMP test accounts for both induced variance in the event window and non-zero cross correlation. The adjusted BMP test is given as:

$$t_{AB} = \frac{\overline{A}}{s_{\overline{A}}} = \frac{\overline{A}\sqrt{n}}{s_A\sqrt{1+(n-1)\overline{r}}}, \tag{7.4}$$

where $\overline{r}$ is the average sample cross correlation of the estimation window residuals and S_A is the standard deviation given by:

$$S_A = \sqrt{\frac{s^2}{1-\overline{r}}}. \tag{7.5}$$

From Equation 7.4, similarities to the original BMP test, from Equation 7.2, can be drawn. Therefore, the adjusted BMP test can also be written as:

$$t_{AB} = t_B\sqrt{\frac{1-\overline{r}}{1+(n-1)\overline{r}}}, \tag{7.6}$$

where t_B is the original BMP test by Boehmer et al. (1991). In the case of no cross-correlation, meaning r is zero, the equation equals the original BMP test (Boehmer et al., 1991). This gives the adjusted BMP the benefit that it only corrects the test statistic if cross correlation is present. In order to test for cumulative abnormal returns, the original BMP test in Equation 7.2 is modified in the following way: standardized average abnormal returns are replaced with mean standardized cumulative abnormal returns in the numerator, and the standard deviation is replaced with the cross sectional standard deviation in the denominator. The adjusted BMP test is therefore given by:

$$t_{AB} = \frac{\overline{CAR}\sqrt{n}}{S_{CAR}\sqrt{1+(n-1)\overline{r}}}, \tag{7.7}$$

where $\overline{CAR}$ is the mean standardized cumulative abnormal return, n is the sample size, $\overline{r}$ is the average sample cross correlation of the estimation window residuals, and S_{CAR} is the cross sectional standard deviation.

7.1.1 Results Adjusted BMP Test

Table 7.1 presents results for the adjusted BMP test. There are 316 positive, 640 neutral and 322 negative events. The null hypothesis for each event window is that average cumulative abnormal returns ($ACAR$) equals zero.

The ten event windows consists of three main types. Post-event windows are event windows which consists of the ESG Score announcement day, plus day(s) after the ESG Score announcement day. Pre-event windows consists of the event day, plus day(s) before the event day. Combined windows consists of the event day plus day(s) both before and after the event day.

Most combined windows for positive and negative events show some evidence of $ACAR$ different from zero, while none of the combined windows for neutral events presents evidence of $ACAR$ different from zero. The evidence is not present in post-event windows, hence the evidence in combined windows seem to be driven by pre-event windows. For pre-event windows, $ACAR$ different from zero is consistently detected for positive events. There could also be, although less consistent, an effect of $ACAR$ different from zero for negative events. For neutral events, $ACAR$ different from zero could be present leading up to the ESG Score event day, but the effect is weak.

All statistically significant $ACAR$ results, across event windows and event groups, are positive. Results which lack statistical significance also seem to have a positive sign, but their lack of statistical significance means that there is not enough evidence to state that they differ from zero.

Table 7.1: Results for the Adjusted BMP Test.

Panel: A - Positive events

Event window	N	ACAR	Min	Max	Adj.BMP
[-5 , 5]	315	0.0114	-0.2402	0.4383	2.1461**
[-3 , 3]	315	0.0067	-0.1769	0.3398	1.7692*
[-1 , 1]	311	0.0040	-0.1821	0.2207	1.4938
[-5 , 0]	316	0.0105	-0.0659	0.5016	2.4943**
[-3 , 0]	316	0.0077	-0.0612	0.2695	2.5950**
[-1 , 0]	316	0.0057	-0.0477	0.1537	2.3821**
[0 , 1]	310	0.0031	-0.1657	0.2134	1.2170
[0 , 3]	314	0.0037	-0.1964	0.3553	0.8865
[0 , 5]	315	0.0052	-0.2376	0.4359	1.0474
[0 , 10]	315	0.0046	-0.2579	0.3114	1.0191

Panel: B - Negative events

Event window	N	ACAR	Min	Max	Adj.BMP
[-5 , 5]	321	0.0115	-0.1592	0.3966	1.8911*
[-3 , 3]	319	0.0064	-0.1380	0.3446	1.3041
[-1 , 1]	321	0.0060	-0.0751	0.3874	1.8425*
[-5 , 0]	322	0.0090	-0.1140	0.5205	1.8643*
[-3 , 0]	322	0.0059	-0.0928	0.4013	1.6049
[-1 , 0]	322	0.0053	-0.0850	0.2752	2.0978**
[0 , 1]	320	0.0030	-0.0781	0.3747	0.6591
[0 , 3]	318	0.0026	-0.1369	0.2661	0.3426
[0 , 5]	321	0.0047	-0.1543	0.2756	0.8222
[0 , 10]	321	0.0044	-0.2042	0.7076	0.3519

Panel: C - Neutral events

Event window	N	ACAR	Min	Max	Adj.BMP
[-5 , 5]	639	0.0065	-0.2154	0.7416	0.9130
[-3 , 3]	637	0.0052	-0.2937	0.3840	1.1484
[-1 , 1]	637	0.0032	-0.0616	0.2311	1.1635
[-5 , 0]	640	0.0043	-0.2071	0.1993	1.0167
[-3 , 0]	640	0.0050	-0.1817	0.1644	1.7494*
[-1 , 0]	640	0.0037	-0.0439	0.1388	1.9231*
[0 , 1]	635	0.0012	-0.0646	0.2194	-0.2319
[0 , 3]	635	0.0017	-0.1896	0.3039	-0.2078
[0 , 5]	637	0.0038	-0.2042	0.7168	0.1542
[0 , 10]	637	0.0022	-0.2781	0.6977	-0.1762

Table 7.1 summarises the results for the adjusted BMP test (Kolari and Pynnönen, 2010), with a null hypothesis of ACAR = 0. Positive events are defined as a year-on-year ESG Score change of more than 10.61%. Negative events are defined as y/y change of -2.77%. Neutral events are defined as y/y change lower than 10.61% and higher than -2.77%. N: Number of analysed events, ACAR: Average cumulative abnormal return, Min: Lowest observed cumulative abnormal return (CAR), Max: Highest observed CAR, Adj.BMP: t-statistic. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.

For positive events, the evidence of $ACAR$ different from zero is statistically significant at the 5 percent level in event windows [-5,5], [-5,0], [-3,0], and [-1,0]. Evidence of $ACAR$ different from zero in event window [-3,3] is statistically significant at the 10 percent level. It seems to be an effect around the event day, more specifically pre-event, which supports Hypothesis 1. The effect is consistent for all pre-event windows at the 5 percent significance level.

For negative events, the evidence of $ACAR$ different from zero is statistical significant at the 10 percent level in event windows [5,5], [-1,1], and [-5,0]. For event window [-1,0], $ACAR$ different from zero is evident at the 5 percent significance level. It could be an effect around the event day, more specifically pre-event, which supports Hypothesis 2. However, evidence is not found in event window [-3,0], and the evidence in event window [-5,0] is only statistically significant at the 10 percent level. It could be that the pre-event effect is isolated to event window [-1,0], or that the effect is not consistently detected because it is not present for negative events.

For neutral events, $ACAR$ different from zero is evident at the 10 percent significance level for event windows [-3,0] and [-1,0]. Hence, there is little to none evidence of $ACAR$ significantly different from zero for neutral events, which supports Hypothesis 3.

The results from the adjusted BMP test generally show stronger evidence of $ACAR$ different from zero for positive and negative events, compared to neutral events. The effect is isolated to pre-event windows, and consistent at the 5 percent significance level for all pre-event windows for positive events.

7.2 The Generalized Rank T-test

The original rank test (Corrado, 1989) has the fundamental feature of transforming abnormal returns to ranks. It means that for each event, daily observations in the estimation window and the event window are assigned a rank, based on best to worst in terms of magnitude of abnormal returns.

Compared to earlier tests, the original rank test (Corrado, 1989) is more robust to event-induced variance, which is found to be 20-50 percent higher in the event window than in the estimation window (Charest (1978); Mikkelson (1981); Rosenstein and Wyatt (1990)). Earlier tests estimates the standard error solely from the estimation window, which could understate them [20]. The rank test (Corrado, 1989) includes event window observations in standard deviation calculations, which reduces the understating bias in case of event induced variance. Corrado and Zivney (1992) improved the rank test by making the test more robust for event-induced volatility.

The rank test was further developed in Kolari and Pynnonen (2011), where two generalized rank tests are presented. They have several robustness features which proves helpful in this thesis. We will apply the GRANK T-test (hereafter GRANKT), as it is more robust to cross-sectional correlation of returns than the other test (Kolari and Pynnonen, 2011).

The most notable improvement of the GRANKT compared to the rank test (Corrado and Zivney, 1992), is that all event day returns are compressed into one cumulative event day. The cumulative event day is an aggregation of all returns in the event window, assigning them to one time point. Thereafter, cumulative event day return is ranked in comparison to the event's daily estimation window returns. It represent a difference from the rank test of Corrado and Zivney (1992)[21], which ranks returns for each day inside the event window separately. Ranking for each separate day before aggregating the ranks means that information regarding magnitude of returns are lost. For example, if there are 30 percent abnormal returns on day three in the event window, the rank would be the same as 5 percent, or even 1 percent, abnormal returns on another day if it is the highest on that specific day.

Thus, an intuitive feature of the cumulative event day is that it does not matter how abnormal returns are distributed across the event window, as all returns per event are always defined as $L_1 + 1$, which is all estimation window returns plus the return of the cumulative event day.

The GRANKT defines Generalised Standardized Abnormal Returns ($GSAR_{it}$)[22]. In essence, $GSAR_{it}$ defines (i) estimation window returns similar to that of Equation 7.1 in Section 7.1 and (ii) the cumulative event day return with a more complex standardization method.

[20] Which follows the line of argument in Section 6.2.

[21] The rank test Corrado and Zivney (1992) was designed to test a single day for abnormal return, but it is commonly aggregated to test for multi-day event windows. The popular aggregation method is discussed in Cowan (1992) and Campbell and Wesley (1993), which is the implied method in this discussion.

[22] The formal approach to calculating $GSAR_{it}$ is somewhat extensive, which makes it impractical to present in this thesis. The interested reader is therefore referred to Kolari and Pynnonen (2011).

The test statistic of the GRANKT (Kolari and Pynnonen, 2011, p. 956) is presented here. $t = 0$ defines the cumulative event day, which means that $\overline{U}_0$ defines abnormal returns at the cumulative event day. The intuition behind the test statistic is to test whether returns at the cumulative event day are abnormal relative to observations in the estimation window. The GRANKT is given by:

$$t_{rank} = Z\left(\frac{T-2}{T-1-Z^2}\right)^{1/2}, \tag{7.8}$$

where

$$Z = \frac{\overline{U_0}}{S_{\overline{U}}}, \tag{7.9}$$

where $\overline{U_0}$ is the mean of U_t, and $S_{\overline{U}}$ and $\overline{U_t}$ are given by:

$$S_{\overline{U}} = \sqrt{\frac{1}{T}\sum_{t\in T}\frac{n_t}{n}\overline{U}_t^2}, \qquad \overline{U}_t = \frac{1}{n_t}\sum_{i=1}^{n_t} U_{it},$$

where n_t is the number of available $GSAR_it$ at time t, $T = T_1 - T_0 + 1$ is the number of observations, and $t \in \mathcal{T} = \{T_0 + 1, .., T_1, 0\}$ is the set of time indices including the estimation period $t = \{T_0 + 1, ..., T_1\}$ and the cumulative abnormal return day at $t = 0$.

7.2.1 Results GRANKT Test

Table 7.2 presents results for the GRANKT, with a null hypothesis of $ACAR$ being equal to zero.

The results show a trend similar to the results for the adjusted BMP test. For combined windows, one event window presents evidence of $ACAR$ different from zero for positive events, and two event windows present evidence of $ACAR$ different from zero for negative events.

For post-event windows, there is no evidence of $ACAR$ for positive events. This also applies to neutral events, although one post-event window shows evidence of $ACAR$ different from zero at the 10 percent significance level. For negative events, one post-event window yields $ACAR$ different from zero, statistically significant at the 5 percent level. This indicates a possible post-event effect, which is primarily driven by abnormal returns on day four and five post-event for negative events. However, this is inconsistent with the findings from the adjusted BMP test.

The evidence of $ACAR$ in combined event windows seems to be driven by evidence in pre-event windows. Both positive and negative pre-event windows show strong evidence of $ACAR$ different from zero, while there is weak to none evidence for neutral events.

Contrary to results for the adjusted BMP test, GRANKT results indicate a stronger effect for negative events in the combined event windows, compared to positive events. However, evidence of $ACAR$ different from zero is strong for both positive and negative events in pre-event windows. This indicates that the effect is weaker post-event day for positive events compared to negative events, although no post-event event windows are statistically significant.

Table 7.2: Results for the Generalized Rank T-test (GRANKT).

Panel: A - Positive events					
Event window	N	ACAR	Min	Max	GRANKT
[-5,5]	315	0.0114	-0.2402	0.4383	1.5004
[-3,3]	315	0.0067	-0.1769	0.3398	2.4176**
[-1,1]	311	0.0040	-0.1821	0.2207	1.4857
[-5,0]	316	0.0105	-0.0659	0.5016	1.4153
[-3,0]	316	0.0077	-0.0612	0.2695	2.9044***
[-1,0]	316	0.0057	-0.0477	0.1537	3.1446***
[0,1]	310	0.0031	-0.1657	0.2134	1.4967
[0,3]	314	0.0037	-0.1964	0.3553	0.7323
[0,5]	315	0.0052	-0.2376	0.4359	1.2798
[0,10]	315	0.0046	-0.2579	0.3114	1.5746
Panel: B - Negative events					
Event window	N	ACAR	Min	Max	GRANKT
[-5,5]	321	0.0115	-0.1592	0.3966	2.9962***
[-3,3]	319	0.0064	-0.1380	0.3446	1.3998
[-1,1]	321	0.0060	-0.0751	0.3874	2.5557**
[-5,0]	322	0.0090	-0.1140	0.5205	3.2648***
[-3,0]	322	0.0059	-0.0928	0.4013	2.8375***
[-1,0]	322	0.0053	-0.0850	0.2752	1.6847*
[0,1]	320	0.0030	-0.0781	0.3747	0.9855
[0,3]	318	0.0026	-0.1369	0.2661	0.5441
[0,5]	321	0.0047	-0.1543	0.2756	1.9840**
[0,10]	321	0.0044	-0.2042	0.7076	0.8694
Panel: C - Neutral events					
Event window	N	ACAR	Min	Max	GRANKT
[-5,5]	639	0.0065	-0.2154	0.7416	1.4821
[-3,3]	637	0.0052	-0.2937	0.3840	-0.0494
[-1,1]	637	0.0032	-0.0616	0.2311	1.0173
[-5,0]	640	0.0043	-0.2071	0.1993	1.1065
[-3,0]	640	0.0050	-0.1817	0.1644	1.6939*
[-1,0]	640	0.0037	-0.0439	0.1388	1.1376
[0,1]	635	0.0012	-0.0646	0.2194	-0.0750
[0,3]	635	0.0017	-0.1896	0.3039	-1.9289*
[0,5]	637	0.0038	-0.2042	0.7168	-0.4780
[0,10]	637	0.0022	-0.2781	0.6977	-0.9883

Table 7.2 summarises the results for the GRANKT, with the null hypothesis of ACAR =0. Positive upgrades are defined as a year-on-year ESG Score change of more than 10.61%. Negative events are defined as y/y change of -2.77%. Neutral events are defined as y/y change lower than 10.61% and higher than -2.77%. Average cumulative abnormal returns (ACAR) are associated with the respective date of the ESG score. N: Number of analysed events, ACAR: Cumulative average abnormal return, Min: Lowest observed cumulative abnormal return (CAR), Max: Highest observed CAR. GRANKT: t-statistic. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.

For positive events, there is evidence of $ACAR$ different from zero in event window [-3,3], statistically significant at the 5 percent level. Event windows [-3,0] and [-1,0] both show evidence of $ACAR$ different from zero, statistically significant at the 1 percent level. Hence, the results for positive events show strong evidence of $ACAR$ different from zero in days prior to the ESG Score event day, which supports Hypothesis 1.

For negative events, evidence of $ACAR$ different from zero is present in event window [-5,5], [-5,0], and [-3,0], statistically significant at the 1 percent level. Furthermore, event window [-1,1] and event window [0,5] both show evidence of $ACAR$ different from zero, statistically significant at the 5 percent level. For event window [-1,0], evidence of $ACAR$ different from zero is statistically significant at the 10 percent level. Results for negative events thus also show strong evidence of $ACAR$ different from zero in days prior to the ESG Score event day, but evidence is weaker closer to the event day. Evidence of $ACAR$ leading up to the event day is in line with Hypothesis 2.

For neutral events, event window [-3,0] and [0,3] presents weak evidence of $ACAR$ different from zero, as they both are statistically significant at the 10 percent level. Generally, there is little evidence of $ACAR$ different from zero for neutral events, which is in line with Hypothesis 3.

7.3 Discussion of Overall Test Results

Overall, results from both tests draw a picture of positive $ACAR$ in pre-event windows for both positive and negative events, indicating positive abnormal price reactions for positive and negative events.

In pre-event windows, results of both tests indicate evidence of positive abnormal price reactions in event window [-1,0] and [-3,0] for positive events. For negative events, results of both tests present evidence of positive abnormal price reactions in event windows [-1,0] and [-5,0]. These event windows are therefore of particular interest, but as they are all pre-event windows, the following discussion is centered around evident effects prior to the event day.

The main argument for why positive events could lead to positive abnormal price reactions is that it could increase investors' future profit[23] expectations. The positive relationship

[23]Future profit expectations are future expected benefit minus future expected costs.

between positive events and stock price reactions are similar to the findings of Bragdon and Marlin (1972), Elayan et al. (2014), and Brogi and Lagasio (2019).

A positive event indicates that a firm has either improved its ESG initiatives, improved its ESG reporting, or both. Either way, it shows an increased focus on ESG initiatives. In turn, it could make investors view the firm as better positioned for the future. As mentioned in Section 1, Nordic countries are aiming to become the world's most sustainable region by 2030, which likely will affect the way firms operate. It is natural to assume that if Nordic countries are to achieve their sustainability goals, public regulators will make it relatively more costly to operate non-sustainable, compared to operating sustainable. For example, public regulators may affect firms by implementing favorable tax incentives for sustainable firms, or fine non-sustainable firms. Investors may therefore view a positive event as a sign of the firm becoming more robust to the sustainable transition.

An additional explanation for positive abnormal price reactions for positive events is an expectation of higher future demand of the stock. It may be based on a notion that more investors, fund managers in particular, are expected to include ESG evaluation as a part of their investment strategies. Increased focus on ESG among investors typically leads to more ESG-friendly fund strategies, which in turn means that fund managers will buy stocks that satisfy ESG evaluation criteria. Hence, a positive event means that the firm is more likely to satisfy such evaluation, leading to a higher future demand of the firm's stock, which should give a positive abnormal price reaction.

Similar to positive events, the main argument for why negative events could lead to positive abnormal price reactions, is increased investor expectations of future profits. A negative event indicate a decreased focus on ESG initiatives, which could increase future profit expectations, depending on investors' views on both the firm and ESG initiatives in general. It is well-known that investor preferences are differentiated, and views on ESG initiatives could also be differentiated. Intuition for a negative event leading to increased profit expectations is that spending less money on ESG initiatives translates to more money spendable in other ways - for example to other growth opportunities or paying dividends. Negative events being correlated with positive abnormal price reactions differs from the findings of Capelle-Blancard and Petit (2017) and Krüger (2015), but similar to the findings of Laplante and Lanoie (1994).

If one group of investors believe ESG initiatives are necessary to accomplish higher future profits, while another group of investors believe ESG initiatives to be an unnecessary cost, it is natural that they react differently to ESG Score announcements. Positive abnormal price reactions for both positive and negative events are therefore plausible.

As mentioned, no post-event windows show evidence of abnormal price reactions. Furthermore, there is little evidence of abnormal price reactions for neutral events. Two main reasons for no abnormal price reactions are that (i) information provided in the ESG Scores is already interpreted by investors, and (ii) information provided in the ESG Scores does not affect stock prices, or at least not enough to detect abnormal price reactions.

In line with (i), the ESG Score is based on public information, as mentioned in Section 1.1. Therefore, investors may already have interpreted firms' ESG initiatives, either through the public information themselves or through another ESG metric. Metrics of ESG initiatives are, as mentioned in Section 1.1, provided by many rating agencies. Other agencies may publish their score closer to the original information disclosure, which would decrease the impact of the ESG Score.

In line with (ii), there are reasons for ESG Scores to not lead to abnormal price reactions. One reason is that firms report their own ESG initiatives, meaning there are likely to be individual differences in reporting. Consequently, the information might not be useful for investors. Another reason is that an ESG Score is not assigned to all firms, thus making it harder for investors to compare a firm's ESG Score to its peers.

The clearest evidence of abnormal price reactions is in the pre-event windows, which indicates that information on ESG Scores might be leaked[24] prior to the announcement. Investors who rely on the ESG Score in their investment decisions would likely be eager to get ahead of the market by getting access to ESG Scores before the announcement.

Another line of reasoning for the presence of abnormal price reactions in days prior to the event day is related to co-founding events. Clustering is already presented in Section 6.1, and many large clusters of ESG Score announcements are located at the last day of the year. It means that days prior to the event day are actually the last trading days of the year, which are somewhat special in finance. It is a known phenomena, which also were discussed in Section 6.1, that investors implement strategic allocations at the end of the

[24]Potential reasons for leaked information were introduced in Section 4.2.

year for tax purposes. This, or other co-founding systematic[25] events, could mean that the abnormal price reactions are caused by something else than ESG Score announcements.

There is little to no evidence of abnormal price reactions for neutral events. The evidence from our tests could be explained by co-founding events which occurs for positive and negative events, but not for neutral events. However, the fact that there is little evidence of abnormal price reactions for neutral events, also makes it more plausible that abnormal price reactions for positive and negative events are correlated to the ESG Score announcements.

7.3.1 Differences in Test Results and Robustness of Tests

The GRANKT tends to reject the null hypothesis at a higher rate than the adjusted BMP test, which could be explained by the adjusted BMP test being more robust for cross-sectional correlation (Kolari and Pynnonen, 2011).

As discussed in Section 6.1, if a test is not sufficiently robust to cross-correlation, it could lead to type 1 errors. Event dates in the data are clustered, as presented in Section 6.1, and it follows that the GRANKT should reject the null hypothesis at a higher rate than the adjusted BMP test if it is not robust enough for cross-sectional correlation. Note that both tests in theory are, but Kolari & Pynnönen states that the adjusted BMP test is additionally robust for cross-sectional correlation. From our test results, the GRANKT does in fact reject the null hypothesis at a higher rate than the BMP test. However, overall test results are strengthened by the fact that both tests draw the same picture, augmenting the results presented in this thesis by showing they are not a result of a shortcoming in a chosen test.

Furthermore, we applied the original BMP test by Boehmer et al. (1991) and the rank test by Corrado and Zivney (1992) to further strengthen the results with respect to robustness of our tests.

Deploying the original BMP test is an interesting check for robustness as the adjusted BMP test by Kolari and Pynnönen (2010) is designed as an improvement of it. As the data is clustered, the original BMP test which assumes independence was expected to

[25] Systematic is meant as re-occurring in the same period as the ESG Scores over the time span of interest.

reject the null hypothesis at a higher rate than the adjusted BMP test.

The rank test (Corrado and Zivney, 1992) should reject the null hypothesis at a lower rate than the GRANKT, because the latter is designed to have higher power than the former. The GRANKT, as discussed in Section 7.2, is an improvement of the rank test as it accounts for event induced variance and cross correlation. The original BMP test and the rank test were conducted for the same events over same event windows. The original BMP test rejects the null hypothesis at a higher rate than the adjusted BMP test, while the rank test rejects the null hypothesis at a lower rate than the GRANKT. These findings are in line with the predictions, and illustrates the importance of choosing robust event study tests.

Complete test results for the original BMP test and rank test are presented in Table A2.7 in Appendix A2.

7.4 Firm Characteristics

In this section, we address Hypothesis 4. Our aim is, as evident from Hypothesis 4, to investigate whether certain firm characteristics are associated with stock price reactions for positive or negative events. Correlation between firm characteristics and stock price reactions are investigated by probit regressions.

7.4.1 Probit Regression Models

Empirical studies have shown that several firm characteristics affect future stock prices (Kogan and Papanikolaou, 2013). A probit regression maximises the likelihood of observing a binary outcome of the dependent variable, by using relevant explanatory variables (Wooldridge, 2013). Hence, the regression outputs are used to interpret if explanatory variables increase the likelihood of the outcome.

In our probit regressions, the dependent variable has two possible outcomes: positive cumulative abnormal returns (CAR) (1), and negative cumulative abnormal returns (0). Positive CAR is thus positive abnormal price reactions, and negative CAR is negative abnormal stock price reactions. As outcomes are binary and mutually exclusive, an increased likelihood of positive abnormal price reactions means a decreased likelihood of negative abnormal price reactions, and vice versa.

The dependent variable is a non-linear function of the explanatory variables, and interpretation of coefficients are consequently not straightforward. Coefficients in probit regressions only indicate at what statistical significance level a relationship is likely to exist, and the direction of the relationship. Therefore, marginal effects are estimated. A marginal effect express the change in the likelihood of positive cumulative abnormal returns, given a change in the explanatory variable - all else equal (Wooldridge, 2013).

The strength of probit regressions are tested by a chi-square test. If the chi-square test statistic, presented in the output of our probit regressions, is not significant, the regression does not account for more of the variation in CAR than one would expect by chance.

The probit model is given as:

$$P(Y = 1) = \phi(\beta_0 + \beta_i x_i), \tag{7.11}$$

where P is the response likelihood, Y is the dependent variable, ϕ is the cumulative distribution function of the normal distribution, β_0 is the regression constant, and β_i is the coefficient estimate for explanatory variable x_i. i is the number of explanatory variables in the estimated model.

7.4.2 Explanatory Variables

One characteristic of interest is a firm's size. There are several reasons for price reactions to differ across firms of different size, where one reason is information asymmetry. Investors typically have more information on larger firms than smaller firms, due to more comprehensive coverage by analysts and the media. Consequently, changes in ESG Score could be less anticipated for smaller firms, in turn leading to a more notable price reaction. We found evidence in our event study that a majority of investors[26] view both positive and negative events as good news, as they both yield positive ACAR. Thus, our information asymmetry argument means that both positive and negative events are less anticipated good news for small firms, compared to large firms. Hence, a negative

[26] The overall impact, as shown event study results in Section 7.1.1 and Section 7.2.1, is positive ACAR, but that does not mean that all investors interpret information the same way. Thus, the use of "majority of investors".

relationship is expected between the likelihood of positive abnormal price reactions and size for both positive and negative events.

On the same line of argument, meaning information on larger firms is more available to the public, the opposite relationship between size and price reactions could also be expected for positive and negative events. More available information could mean that large firms face more pressure from public regulators, compared to smaller firms. Given the Nordic goals of sustainability (Nordic Council, 2020), public regulators in the Nordic have an incentive to progress sustainable development. Therefore, large firms may benefit more from positive events compared to smaller firms, as pressure from public regulators are relatively more decreased. It indicates a positive relationship between positive abnormal price reactions and size. Using this argument for negative events, the relationship between positive abnormal price reactions and size should be negative, as we expect public regulators to interpret negative events as bad news.

Dang et al. (2018) find that several measures should be used for size to ensure precise results, as different measures have both advantages and disadvantages. Total assets is an accounting based measure, based on historical information, and reflects the firm's amount of resources. However, the firm's accounting preferences could impact total assets in ways which reduces the precision of comparing total assets across firms. For example, (i) IFRS[27] allows individual variation in estimations, (ii) depreciation might have a time lag in accounting-based measures, and (iii) firms vary in their debt exposure.

Using market capitalisation, a market-based measure, avoids the mentioned problems of total assets. However, stock market imperfections means that market-based measures also have disadvantages. For example, the liquidity of a stock could affect its price, in turns causing stock price an imprecise measure for size.

Total assets does not rely on investor expectations. Therefore, both total assets and market capitalisation are used as proxies for size, and used as explanatory variables in our probit regressions. In order to avoid skewness, the natural logarithm is used[28]. The variables are defined as *ln.asset* and *ln.mcap* in our regressions.

[27] IFRS stands for International Financial Reporting Standards, and firms included in this study follow these.

[28] Market capitalization is fat-tailed, meaning there are more large cap firms than small cap firms. Using natural logarithms are a way of normalising the distribution.

Another characteristic of interest is the firm's initial ESG Score, which is typically set the previous year. A high initial ESG Score reflects that the firm is already doing well in terms of ESG initiatives and reporting. A positive event could ensure investors that the firm remains committed to improving their ESG initiatives. Hence, a positive relationship is expected between the likelihood of positive abnormal price reactions and ESG Score for positive events.

For negative events, a firm with a high ESG Score initially, could be worse of compared to a firm with a low ESG score initially. Intuition is that investors who value ESG initiatives are more likely to invest in firms with a high ESG Score. Firms with a low initial ESG Score, however, were never attractive to investors due to ESG initiatives in the first place. Thus, the likelihood of positive abnormal price reactions is expected to be higher for firms with an initial low ESG Score, compared to firms with an initial high ESG Score. ESG Score is set as the dummy $ESG.d$ in the probit regressions, set to 1 if the firm has a score of 50 or higher, and 0 if not[29].

The third characteristic of interest is related to industry. The motivation for including fossil fuels as an industry dummy is that operations within fossil fuels are by nature not sustainable, and the industry is being criticized for their impact on the environment. Furthermore, the fossil fuels industry is viewed as particularly relevant for this analysis as the Norwegian economy is centered around it. The motivation for including a dummy for banking and investment services is the findings of Brogi and Lagasio (2019), as presented in Section 2. The researchers found that increased ESG initiatives were associated with higher return on assets (ROA) for the banking and investment industry, relative to other industries. If the same relationship exists in the Nordic market, we expect a higher likelihood of positive abnormal price reactions for the banking sector in association with a positive event. Two dummy variables are defined: $Fossil.d$ is set to 1 if the firm operates within fossil fuels, and 0 if not. $Bank.d$ is set to 1 if the firm operates within banking and investments services, and 0 if not.

The fourth characteristic of interest is financial performance. Intuition is that ESG initiatives are costly, and financial performance could indicate how well firms are suited to invest in them. Investors may interpret that financially well-performing firms could afford

[29]The ESG Score range from 1-100. A high score is defined as a ESG Score of 50 or higher

ESG investing to a larger degree than firms which are not performing as well financially. Moreover, investors could find focusing on the underlying operation more sensible than investing in ESG initiatives when financial performance is weak. If so, the relationship between financial performance and the likelihood of positive abnormal price reactions are expected to be positive for positive events and negative for negative events.

Measuring financial performance might be done in various forms. ROA is according to Hult et al. (2008) one of the most commonly used proxies for financial performance in empirical research. ROA is measured as net income/total assets, and defined as *ROA* in our probit regressions. To ensure robustness, two different variables are applied as proxies for financial performance. The second proxy is EBITDA-margin[30], also an accounting measure, but one that limits some of the imprecision mentioned for total assets. For example, EBITDA-margin is not sensitive to choice of depreciation profile and one-off financial effects outside of the operation. Ideally, EBITDA-margin should be normalized to provide a more accurate comparison between firms, but that is outside the scope of this thesis. EBITDA-marign is defined as *Ebitda* in the probit regression.

The fifth characteristic of interest is volatility. Volatility is measured by the explanatory variable *Beta*, which is calculated in the events' respective estimation window. Beta is the volatility of the stock relative to the market, measured by the respective value-weighted market index, as specified in Section 4.4. The motivation for including beta as an explanatory variable is that investors might interpret a firm with a higher *Beta* as riskier, and investor who value ESG initiatives could therefore be more reluctant to invest in these despite a positive event. For negative events, investors who value ESG initiatives could view firms with a low *Beta* as more robust to ESG Score changes as they by nature are less risky. Thus, we expect the relationship between the likelihood of positive abnormal price reactions and beta to be negative for positive events and negative events.

The last characteristic of interest is time. A time dummy is included, which corresponds to 1 if the event occurs in 2017 or later, and 0 otherwise. The dummy is included because the focus on ESG initiatives in investment decisions has undoubtedly increased over

[30]EBITDA-margin is EBITDA as a percentage of sales. EBITDA is Earnings Before Interest, Taxes, Depreciation and Amortization. Note that the margin typically varies across industries, but it does not affect the purpose of the variable in this analysis.

time, as mentioned Section 1. We therefore expect investors to be more positive to ESG initiatives, and thus more positive to positive events and less positive to negative events, when the time dummy equals one. Therefore, we expect a positive relationship between the likelihood of positive abnormal price reactions and the time dummy for positive events. The relationship between the likelihood of positive abnormal price reactions and the time dummy is expected to be negative for negative events, and the time dummy is defined as $Time.d$ in the probit regressions.

To isolate the relationship between the likelihood of positive abnormal price reactions and each explanatory variable, correlation between explanatory variables is examined. Results are presented in a correlation matrix, which is found in Figure A3.1 in Appendix A3. The only variables which are strongly correlated are $ln.asset$ and $ln.mcap$, and they are therefore not included in the same models[31].

Thus, our probit models are given in the following. Model 1 and model 2 differs by the former including $ln.asset$ as an explanatory variable, while it is replaced by $ln.mcap$ in the latter. Model 3 is structured the same way as model 1, but applied to another event window. Likewise, model 4 is structured the same way as model 2, but applied to another event window. Model 1 is given in Equation 7.12, and model 2 is given in Equation 7.13.

$$\begin{split} P(PositiveCAR = 1) = \phi(&\beta_0 + \beta_1 Beta_i + \beta_2 ROA_i + \beta_3 Ebitda_i + \beta_4 ln.asset_i \\ &+ \beta_5 ESG.d_i + \beta_6 Fossil.d_i + \beta_7 Bank.d_i + \beta_8 Time.d_i)\,. \end{split} \tag{7.12}$$

$$\begin{split} P(PositiveCAR = 1) = \phi(&\beta_0 + \beta_1 Beta_i + \beta_2 ROA_i + \beta_3 Ebitda_i + \beta_4 ln.mcap_i \\ &+ \beta_5 ESG.d_i + \beta_6 Fossil.d_i + \beta_7 Bank.d_i + \beta_8 Time.d_i)\,. \end{split} \tag{7.13}$$

7.4.3 Probit Regression Output

Following our discussion in Section 7.3, event windows [-1,0] and [-3,0] for positive events, and event windows [-1,0] and [-5,0] for negative events are chosen as most relevant for further analysis. Hence, the effect of firm characteristics are examined in these windows.

Table 7.3 presents probit regression results for positive events. Size proxies $ln.asset$ and $ln.mcap$ are significant in all four regression models. Coefficients for both variables are

[31] It follows from correlation results and intuition that including both size variables in the same model could cause multicollinearity problems.

negative, meaning larger firms are less likely to have positive abnormal price reactions for positive events. Consequently, there is a positive relationship between being a small firm[32] and the likelihood of positive abnormal price reactions after a positive event. A possible reason for the negative relationship, as introduced in Section 7.4.2, is that investors are more surprised by positive events for small firms due to information asymmetry.

Our choice of including two proxies for size, in line with Dang et al. (2018), means that our results are robust as coefficients of both proxies are pointing in the same direction. There seems to be a relationship between size and the likelihood of positive abnormal price reactions.

As *Bank.d* is not statistically significant, we do not find evidence to support the findings of Brogi and Lagasio (2019). A possible explanation is that a broader focus on ESG across industries in the Nordic implies that the banking industry does not stand out in the same way as in the United States. Similar to the banking industry, *Fossil.d* indicates that the fossil fuels industry does not seem to stand out relatively to other industries.

Neither *ROA* nor *Ebitda* present evidence of a relationship between financial performance and the likelihood of observing positive abnormal price reactions for positive events. It could mean that investors do not have a preference between a financially well-performing firm and a financially weak-performing firm when ESG initiatives are notably improved. Another possible explanation for no evident relationship is that our explanatory variables do not represent precise proxies for financial performance.

Coefficients for *ESG.d* do not present evidence of an existing preferences for investors related to the previous year's ESG Score. Furthermore, *Beta* is not statistically significant for any model, indicating that (i) volatility does not affect the likelihood of positive abnormal price reactions for positive events, or (ii) the measurement of volatility does not capture the effect.

[32] Small firms and large firms are interpreted within the data set. A small firm is thus small relatively to other firms examined.

Table 7.3: Probit Regression results for Positive Events.

	Dependent variable (1 = Positive CAR, 0 = Negative CAR)			
	Event window [-1,0]		Event window [-3,0]	
	Model(1)	Model(2)	Model(3)	Model(4)
Beta	-0.1299	-0.0720	0.0184	0.1212
ROA	-0.4957	0.0934	0.0173	0.5890
Ebitda	-0.3020	-0.2173	0.0728	0.1468
ln.asset	-0.1036*	-	-0.1138*	-
ln.mcap	-	-0.1970***	-	-0.2726***
ESG.d	-0.0413	0.0321	-0.0467	0.0841
Fossil.d	0.3399	0.2258	0.4340	0.2636
Bank.d	0.3038	0.1356	-0.0667	-0.2501
Time.d	0.3951***	0.3725**	0.4630***	0.4375***
cons	2.4065*	4.2924***	2.3042*	5.5521***
N	294	294	294	294
Pseudo R2	0.0408	0.0538	0.0504	0.0798
Chi-square	16.32**	21.49***	20.31***	32.11***

Table 7.3 summarises probit regression results for positive events in event window [-1,0] and [-3,0]. The dependent variable is a binary variable with 2 possible outcomes where 1 = Positive cumulative abnormal return (CAR), and 0 = Negative CAR. Positive events are defined as an ESG score year-on-year change $> 10.61\%$. The probit regression is estimated by:
$P(Positive\ CAR = 1) = \phi(\beta_0 + \beta_1 Beta_i + \beta_2 ROAi + \beta_3 Ebitda_i + \beta_4 ln.asset_i + \beta_5 ln.mcap_i + \beta_6 ESG.d_i + \beta_7 Fossil.d_i + \beta_8 Bank.d_i + \beta_9 Time.d_i)$. Beta: Stock volatility (stock return variation in relation to market return variation) in estimation window for firm i. Ebitda: EBITDA margin (Percentage measure calculated by EBITDA/total sales for firm i). ROA: Return on assets (Calculated as income after taxes for fiscal year (FY) divided by average total assets of FY). ln.asset: Natural logarithm of average total assets in FY. ln.mcap: Natural logarithm of market capitalization at the end of FY (Calculated by last share price of FY*total shares outstanding). ESG.d: Dummy variable which equals 1 if ESG scores >50 for previous year, and 0 otherwise. Fossil.d: Dummy variable equal 1 if firm i operates in the fossil fuel sector. Bank.id: Dummy variable equal 1 if firm i operates in the banking sector. Time.d: Time dummy that equals 1 if the observation was in 2017 or later. cons: Regression constant. N: Total number of observations. Pseudo R2: Mc Faddens R2. Model (1): Variable ln.mcap excluded. Model (2): Variable ln.asset excluded. Model (3): Variable ln.mcap excluded. Model (4): Variable ln.asset excluded. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.

The time dummy $Time.d$ is statistically significant at the 1 percent for all models but model 2, where it is statistically significant at the 5 percent level. Therefore, there is strong evidence of positive events dated 2017 or later being more likely to have positive abnormal price reactions in days prior to the event. It is in line with our reasoning in Section 7.4.2, where we predicted this relationship.

Marginal effects for positive and negative events are estimated for all coefficients, and results are presented in Table A3.3 in Appendix A3. Only findings important for our analysis are discussed here.

A 1 percent increase in *ln.mcap* decreases the likelihood of positive cumulative abnormal returns by 7.25 percent for model 2, and 9.83 percent for model 4. For *ln.asset*, an increase of 1 percent in total assets decreases the likelihood of positive cumulative abnormal returns by 3.87 percent in model 1 and 4.24 percent in model 3. If *Time.d* equals 1, meaning the event occurs in 2017 or later, there is a 14.75 percent higher likelihood of positive abnormal price reactions in model 1, and 13.70 percent in model 2. For event window [-3,0], the likelihood of positive abnormal price reactions increases by 17.25 percent for model 3, and 15.77 percent for model 4 when *Time.d* equals 1.

Table 7.4 presents probit regression results for negative events. Similar to the findings for positive events, dummies *Fossil.d* and *Bank.d*, as well as *ESG.d*, do not indicate a relationship with the likelihood of positive abnormal price reactions for negative events.

Results for negative events show a negative relationship between the likelihood of observing positive abnormal price reactions, and the firm's *ROA* for model 1, 3, and 4. An explanation for the negative relationship is that firms with weak financial performance are less suited to implement costly ESG initiatives. Thus, investors would interpret a negative event for financially weak-performing firms as better news relatively to financially well-performing firm. On the other hand, *Ebitda* supports no such relationship, and *ROA* is only highly significant for event window [-1,0]. The possible relationship between financial performance and the likelihood of positive abnormal price reactions are therefore interpreted with caution.

Results for *Beta* indicates little evidence of a relationship between volatility and the likelihood of abnormal stock price reactions. Model 3 and 4 indicate a negative relationship between the likelihood of positive abnormal price reactions and volatility. However, results are only significant at the 10 percent level, and only for event window [-5,0].

Furthermore, size proxies *ln.asset* and *ln.mcap*, are significant at the 1 percent in model 3 and model 2. In model 4, *ln.mcap* is statistical significant at the 10 percent level. In model 1, *ln.asset* is statistically significant at the 10 percent level. Similar to results for positive events, coefficients for *ln.asset* and *ln.mcap* are negative, meaning larger firms are less likely to have positive abnormal price reactions for negative events.

Table 7.4: Probit Regression Results for Negative Events.

	Dependent variable (1 = Positive CAR, 0 = Negative CAR)			
	Event window [-1,0]		Event window [-5,0]	
	Model(1)	Model(2)	Model(3)	Model(4)
Beta	-0.0794	-0.0570	-0.3594*	-0.3724*
ROA	-2.9892***	-1.9808**	-1.3755*	-0.6989
Ebitda	0.4610	0.4039	-0.1176	-0.1233
ln.asset	-0.1533**	-	-0.1530***	-
ln.mcap	-	-0.1938***	-	-0.1103*
ESG.d	-0.2409	-0.2162	-0.1969	-0.2504
Fossil.d	-0.1201	-0.1286	0.1268	0.1289
Bank.d	0.4801	0.2525	0.3046	-0.0341
Time.d	0.4517***	0.4775***	0.0227	-0.0681
cons	3.4924***	4.2563***	4.0693***	3.1023**
N	315	315	315	315
Pseudo R2	0.0895	0.0931	0.0549	0.0455
Chi-square	39.08***	40.63***	23.72***	19.65***

Table 7.4 summarises probit regression results for positive events in event window [-1,0] and [-3,0]. The dependent variable is a binary variable with 2 possible outcomes where 1 = Positive cumulative abnormal return (CAR), and 0 = Negative CAR. Negative events are identified with y/y change of -2.77%. The probit regression is estimated by:
$P(Positive\ CAR = 1) = \phi(\beta_0 + \beta_1 Beta_i + \beta_2 ROAi + \beta_3 Ebitdai + \beta_4 ln.asset_i + \beta_5 ln.mcap_i + \beta_6 ESG.d_i + \beta_7 Fossil.d_i + \beta_8 Bank.d_i + \beta_9 Time.d_i)$. Beta: Stock volatility (stock return variation in relation to market return variation) in estimation window for firm i. Ebitda: EBITDA margin (Percentage measure calculated by EBITDA/total sales for firm i). ROA: Return on assets (Calculated as income after taxes for fiscal year (FY) divided by average total assets of FY). ln.asset: Natural logarithm of average total assets in FY. ln.mcap: Natural logarithm of market capitalization at the end of FY (Calculated by last share price of FY*total shares outstanding). ESG.d: Dummy variable which equals 1 if ESG scores >50 for previous year, 0 otherwise. Fossil.d: Dummy variable equal 1 if firm i operates in the fossil fuel sector. Bank.id: Dummy variable equal 1 if firm i operates in the banking sector. Time.d: Time dummy that equals 1 if the observation was in 2017 or later. cons: Regression constant. N: Total number of observations. Pseudo R2: Mc Faddens R2. Model (1): Variable Ebitda and ln.mcap excluded. Model (2): Variable Ebitda and ln.asset excluded. Model (3): Variables ROA and ln.mcap excluded. Model (4): Variables ROA and ln.asset excluded. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.

The time dummy $Time.d$ is statistically significant at the 1 percent level for model 1 and 2, and not statistically significant at all for model 3 and 4. Hence, it seems to be a positive relationship between the likelihood of positive abnormal price reactions and $Time.d$ close to a negative event, but the effect is not present when examining the longer event window [-5,0]. The relationship is positive, meaning positive abnormal price reactions is more likely for events in 2017 or later. It is in opposition of our expected relationship, but in line with the results of the event study - where we found negative news to be good news for investors. Results for $Time.d$ in event window [-1,0] indicates that this effect is stronger for recent years.

From Table A3.3 in Appendix A3, the following marginal effects are estimated. A 1 percent increase in $ln.mcap$ decreases the likelihood of positive abnormal price reactions by 6.95 percent for model 2, and 4.14 percent for model 4. For $ln.asset$, an increase of 1 percent in total assets decreases the likelihood of positive abnormal price reactions by 5.52 percent in model 1 and 5.69 percent in model 3. For $Time.d$, negative events in 2017 and later have a 16.28 percent higher likelihood of positive abnormal price reactions in event window [-1,0] for model 1. The likelihood of positive abnormal price reactions is 17.13 percent when $ln.asset$ is replaced by $ln.mcap$.

To summarize, results for both positive and negative events indicates that size and $Time.d$ interacts with the likelihood of positive abnormal price reactions in our selected event windows. For size, smaller firms have a higher likelihood of positive abnormal price reactions, an effect which seems to be present for all event windows, where $ln.mcap$ seems to be a better explanatory variable than $ln.asset$. For $Time.d$, more recent events are more likely to have positive abnormal price reactions, an effect which is highly statistically significant for both event windows for positive events, and also highly statistically significant for event window [-1,0] for negative events. The effect is not present for the longer pre-event window for negative events, [-5,0].

In addition to our analysis on firm characteristics correlated to stock price reactions around positive and negative events, we have conducted a multionominal logistic regression to examine which firm characteristics correlate to firms earning a positive or a negative event in the first place. Results, as well as a short explanation of the multinominal logistic regression, is placed in Appendix A4. The interested reader is therefore referred to Section A4.1 in Appendix A4.

8 Conclusion

Previous research on the relationship between ESG related announcements and stock price reactions provide no clear consensus on an existing relationship between them. Furthermore, there is reason to believe that stock price reactions to ESG announcements differ across regions, based on investor preferences.

We examine stock price reactions to ESG Score announcements in the Nordic market between 2011 and 2019, where the Nordic market is defined as Sweden, Denmark, Norway, and Finland. These countries are of particular interest due to their focus on sustainability, evident from the goal of becoming the most sustainable region in the world by 2030 (Nordic Council, 2020).

Our primary aim was to examine the most extreme announcements, which is defined by how much the new ESG Score differ from the last year's ESG Score. Therefore, year-on-year changes were divided into positive, negative, and neutral events. Positive events are defined as ESG Score changes of more than 10.61 percent. Negative events are defined as ESG Score changes of less than -2.77, while neutral events are changes in between positive and negative events.

Our first hypothesis was that announcements of positive events are associated with abnormal stock price reactions, while the second hypothesis was that negative events are associated with abnormal stock price reactions. There is evidence of abnormal stock price reactions around the ESG Score announcements for both positive and negative events. Evidence for both event groups is located in pre-event windows, which could indicate that information regarding ESG Scores is leaked prior to the event date. Results of the adjusted BMP test detects abnormal price reactions primarily for positive events, while the results of the GRANKT test shows the strongest evidence for negative events. However, the results of both tests present evidence of abnormal price reactions for both positive and negative events. These findings indicate that the ESG Score announcements represent new information for investors.

Our third hypothesis was that neutral events are not associated with abnormal stock price reactions. There is no evidence of abnormal stock price reactions around neutral events at the 5 percent statistical significance level.

Abnormal price reactions are positive for both positive and negative events, indicating that both positive and negative events could present good news for investors. These findings point to investor views being differentiated, where some investors value ESG initiatives and others do not.

As most events are clustered around the turn of a new year, there could be systematic co-founding events which causes abnormal price reactions. However, that does not explain different findings for positive and negative events, compared to neutral events. On the contrary, it supports a notion that evidence of abnormal price reactions could be related to the ESG Score announcements.

Our fourth and final hypothesis was that certain firm characteristics are correlated with positive and/or negative stock price reactions. On that note, evidence shows that smaller firms are more likely to be correlated with positive abnormal stock price reactions for both positive and negative events. Furthermore, positive abnormal stock price reactions are more likely prior to the event day if the event occurred in 2017 or later. A possible reason for the findings across time, is that in later years, ESG initiatives are included in more investment decisions. Thus, as we found in the event study that both positive and negative events represents good news to investors, it also makes sense that the relationship between the likelihood of positive abnormal price reactions and our time dummy is positive. For the findings on size, an explanation is information asymmetry. There is typically less available information on small firms compared to large firms, which means events are less anticipated by investors. Again, as both positive and negative events are good news to investors, it is reasonable that smaller firms are more likely to be associated with positive abnormal stock price reactions for both event types.

As more companies get rated by Thomson Reuters, their database on ESG Scores are likely to become more extensive over time. Hence, examining stock price reactions to ESG Score announcements could be even more interesting in the future, as it would be possible to draw conclusions for a larger part of the Nordic stock universe.

Comparing price reactions across more countries could also be interesting, in order to examine the differences between investors' view on ESG initiatives in detail. Furthermore, future researchers are encouraged to investigate additional firm characteristics which could be correlated with positive stock price reactions around ESG Score announcements.

References

Aktas, N., Bodt, E., and Cousin, J.-G. (2007). Event studies with a contaminated estimation period. *Journal of Corporate Finance*, 13:129–145.

Alexander, G. and Buchholz, R. (1978). Corporate Social Responsibility and Stock Market Performance. *The Academy of Management Journal*, 21:479–486.

Alfred Berg (2020). Nordic sustainable and socially responsible investing - an old habit in a strong esg environment. Retrieved 12/12-20 from https://www.alfredberg.com/nordic-sustainable-and-socially-responsible-investing-an-old-habit-in-a-strong-esg-environment/.

Ball, R. and Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research*, 6(2):159–178. Publisher: [Accounting Research Center, Booth School of Business, University of Chicago, Wiley].

Binder, J. J. (1998). The Event Study Methodology Since 1969. *Review of Quantitative Finance & Accounting*, 11(2):111–137. Publisher: Springer Nature.

Boehmer, E., Masumeci, J., and Poulsen, A. B. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics*, 30(2):253–272.

Borooah, V. K. (2002). *Logit and Probit: Ordered and Multinominal Models*. Sage Publications, Inc., Thousand Oaks, CA.

Bragdon, J. H. and Marlin, J. A. (1972). Is Pollution Profitable? *Risk Management*, pages 9–18.

Brogi, M. and Lagasio, V. (2019). Environmental, social, and governance and company profitability: Are financial intermediaries different? *Corporate Social Responsibility and Environmental Management*, 26(3):576–587.

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.

Campbell, C. J., Cowan, A. R., and Salotti, V. (2010). Multi-country event-study methods. *Journal of Banking & Finance*, 34(12):3078–3090.

Campbell, C. J. and Wesley, C. E. (1993). Measuring security price performance using daily NASDAQ returns. *Journal of Financial Economics*, 33(1):73–92.

Capelle-Blancard, G. and Petit, A. (2017). Every Little Helps? ESG News and Stock Market Reaction. *Journal of Business Ethics*, 157(2):543–565. Publisher: Springer Nature.

Charest, G. (1978). Split information, stock returns and market efficiency-I. *Journal of Financial Economics*, 6(2-3):265–296.

Cheng, B., Ioannou, I., and Serafeim, G. (2014). Corporate social responsibility and access to finance. *Strategic Management Journal*, 35(1):1–23. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/smj.2131.

Cheung, A. (2011). Do Stock Investors Value Corporate Sustainability? Evidence from an Event Study. *Journal of Business Ethics*, 99:145–165.

Chordia, T., Roll, R., and Subrahmanyam, A. (2005). Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics*, 76(2):271–292.

Corrado, C. J. (1989). A nonparametric test for abnormal security-price performance in event studies. *Journal of Financial Economics*, 23(2):385–395.

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. *Journal of Financial & Quantitative Analysis*, 27(3):465–478. Publisher: Cambridge University Press.

Cowan, A. R. (1992). Nonparametric Event Study Tests. *Review of Quantitative Finance & Accounting*, 2(4):343–358. Publisher: Springer Nature.

Dang, C., (Frank) Li, Z., and Yang, C. (2018). Measuring firm size in empirical corporate finance. *Journal of Banking & Finance*, 86:159–176.

Drempetic, S., Klein, C., and Zwergel, B. (2020). The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review. *Journal of Business Ethics*, 167(2):333–360.

Elayan, F., Li, J., Liu, Z., Meyer, T., and Felton, S. (2014). Changes in the Covalence Ethical Quote, Financial Performance and Financial Reporting Quality. *Journal of Business Ethics*, 134(3):369–395. Publisher: Springer Nature.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2):383–417. Publisher: [American Finance Association, Wiley].

Fama, E. F. (1991). Efficient Capital Markets: II. *The Journal of Finance*, 46(5):1575–1617. Publisher: [American Finance Association, Wiley].

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3):283–306.

Fama, E. F., Fisher, L., Jensen, M. C., and Roll, R. (1969). The Adjustment of Stock Prices to New Information. *International Economic Review*, 10(1):1. Publisher: Wiley-Blackwell.

Harrington, S. E. and Shrider, D. G. (2007). All Events Induce Variance: Analyzing Abnormal Returns When Effects Vary across Firms. *Journal of Financial & Quantitative Analysis*, 42(1):229–256. Publisher: Cambridge University Press.

Hillmer, S. C. and Yu, P. L. (1979). The market speed of adjustment to new information. *Journal of Financial Economics*, 7(4):321–345.

Hult, G. T. M., Ketchen, D. J., Griffith, D. A., Chabowski, B. R., Hamman, M. K., Dykes, B. J., Pollitte, W. A., and Cavusgil, S. T. (2008). An assessment of the measurement of performance in international business research. *Journal of International Business Studies*, 39(6):1064–1080.

Ioannou, I. and Serafeim, G. (2010). What Drives Corporate Social Performance? The Role of Nation-level Institutions. *Journal of International Business Studies*, 43.

Jaffe, J. P. (1974). Special Information and Insider Trading. *Journal of Business*, 47(3):410–428. Publisher: University of Chicago Press.

Kogan, L. and Papanikolaou, D. (2013). Firm Characteristics and Stock Returns: The Role of Investment-Specific Shocks. *The Review of Financial Studies*, 26(11):2718–2759. Publisher: [Oxford University Press, Society for Financial Studies].

Kolari, J. W. and Pynnonen, S. (2011). Nonparametric rank tests for event studies. *Journal of Empirical Finance*, 18(5):953–971.

Kolari, J. W. and Pynnönen, S. (2010). Event Study Testing with Cross-sectional Correlation of Abnormal Returns. *The Review of Financial Studies*, 23(11):3996–4025. Publisher: [Oxford University Press, Society for Financial Studies].

Kothari, S. and Warner, J. B. (2007). Econometrics of Event Studies. In *Handbook of Empirical Corporate Finance SET*, pages 3–36. Elsevier Science & Technology.

Kothari, S. P. and Warner, J. B. (1997). Measuring long-horizon security price performance. *Journal of Financial Economics*, 43(3):301–339.

Krivin, D., Patton, R., Rose, E., and Tabak, D. (2003). Determination of the Appropriate Event Window Length in Individual Stock Event Studies. *SSRN Electronic Journal*.

Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of Financial Economics*, 115(2):304–329.

Laplante, B. and Lanoie, P. (1994). The market response to environmental incidents in Canada: A theoretical and empirical analysis. *Southern Economic Journal*, 60(3):657. Publisher: John Wiley & Sons, Inc.

Lev, B. (1989). On the Usefulness of Earnings and Earnings Research: Lessons and Directions from Two Decades of Empirical Research. *Journal of Accounting Research*, 27:153–192. Publisher: [Accounting Research Center, Booth School of Business, University of Chicago, Wiley].

Lyon, J. D., Barber, B. M., and Tsai, C.-L. (1999). Improved Methods for Tests of Long-Run Abnormal Stock Returns. *The Journal of Finance*, 54(1):165–201. Publisher: [American Finance Association, Wiley].

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

Maynes, E. and Rumsey, J. (1993). Conducting event studies with thinly traded stocks. *Journal of Banking & Finance*, 17(1):145–157.

Mikkelson, W. H. (1981). Convertible calls and security returns. *Journal of Financial Economics*, 9(3):237–264.

Nasdaq Nordic (2019). Market Cap Segment Review at Nasdaq Nordic Exchanges. Retrieved 27/10-20 from https://www.nasdaq.com/press-release/market-cap-segment-review-at-nasdaq-nordic-exchanges-2019-12-18.

Nordic Council (2020). About the Nordic Council of Ministers. Retrieved 28/11-20 from: https://www.norden.org/en/information/about-nordic-council-minister.

Oberndorfer, U., Schmidt, P., Wagner, M., and Ziegler, A. (2011). Does the stock market value the inclusion in a sustainability stock index? An event study analysis for German firms. *Journal of Environmental Economics and Management*, 66(3):497–509.

Oslo Børs (2020). Oslo Børs MidCap Index. Retrieved 27/11-20 from: http://live.euronext.com/nb/product/indices/NO0010735640-XOSL/overview.

Patell, J. M. (1976). Corporate Forecasts of Earnings Per Share and Stock Price Behavior: Empirical Test. *Journal of Accounting Research*, 14(2):246.

Plantinga, A., Scholtens, B., and Duuren, E. (2015). ESG Integration and the Investment Management Process: Fundamental Investing Reinvented. *Journal of Business Ethics*, 138.

PricewaterhouseCoopers (2020). 2022 - The growth opportunity of the century. Retrieved 26/11-20 from: https://www.pwc.lu/en/sustainable-finance/esg-report-the-growth-opportunity-of-the-century.html.

Refinitiv (2020). Environmental social and governance (ESG) Scores from Refinitiv. Retrieved 28/10-20 from: https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/esg-scores-methodology.pdf.

Rosenstein, S. and Wyatt, J. G. (1990). Outside directors, board independence, and shareholder wealth. *Journal of Financial Economics*, 26(2):175–191.

Salinger, M. (1992). Standard Errors in Event Studies. *The Journal of Financial and Quantitative Analysis*, 27(1):39.

Sinani, E., Stafsudd, A., Thomsen, S., Edling, C., and Randøy, T. (2008). Corporate governance in Scandinavia: comparing networks and formal institutions. *European Management Review*, 5(1):27–40.

Sustainable Insight Capital Management (2016). Who are the ESG rating agencies? Technical report, Sustainable Insight Capital Management.

United Nations (2020). The 17 Goals of Sustainable Development. Retrieved 28/11-20 from: https:// sdgs.un.org/goals.

Vance, S. G. (1975). Are Socially Responsible Corporations Good Investment Risks? *Management Review*, 64(8):18. Publisher: American Management Association.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning, Mason, Ohio, 5th edition.

Appendix

A1 Appendix A

A1.1 Market Model

Figure A1.1: Residual Plots for Axfood AB Listed at the Nasdaq Stockholm Stock Exchange in Sweden.

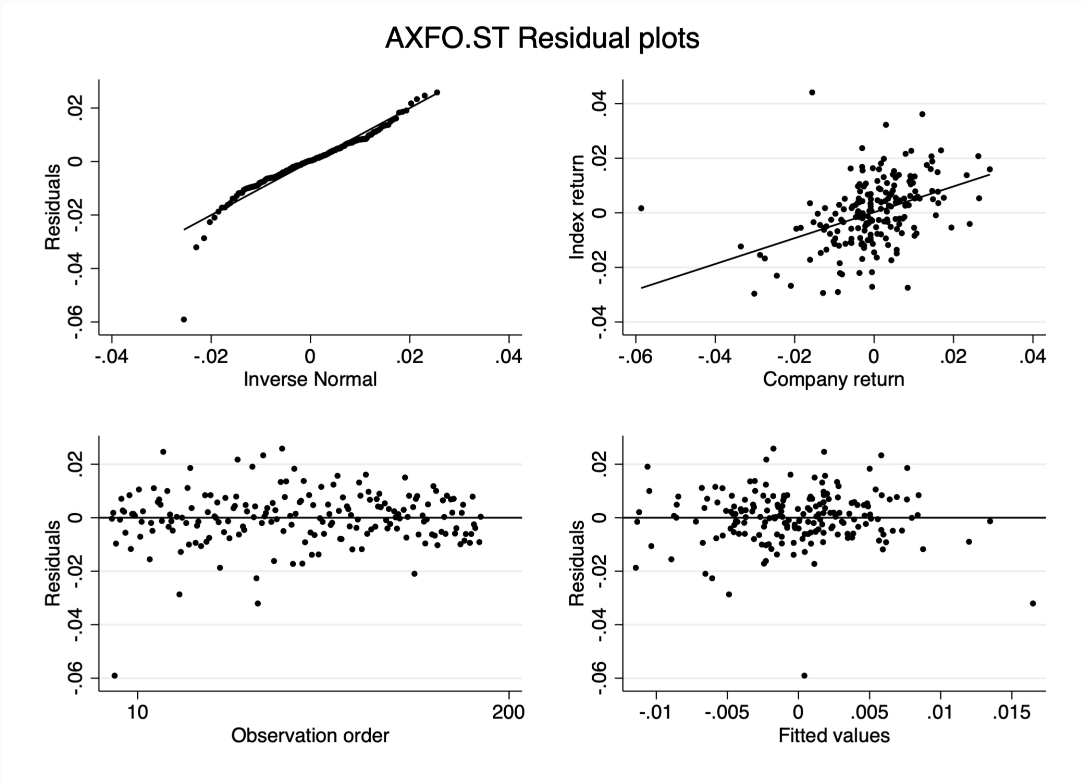


Figure A1.1 shows four different residual plots for Axfood AB regressed on the Value Weighted Index (OMXS30NEXTGI), in the estimation window [-200,-10] with event date 31.12.2012. The upper left plot is the normal probability plot (qnorm), which indicates whether the residuals are normally distributed or not. To the upper right is a linearity plot, we test whether the linearity assumption between firm returns and index returns hold. To the lower left is a residuals vs. observations order plot, which indicates whether the residuals are uncorrelated or not. To the lower right is the residual vs. fitted value plot, which indicates whether the residuals have constant variance (homoscedasticity) or not (heteroscedasticity).

Deviations from the trend line are minor, and the normality assumption is accepted. There are a few outliers in the linearity plot, but observations generally do not deviate much from linearity. No clear relationship between residuals is identified in the residual vs. observation order plot , and we find the correlation assumption reasonable. Lastly, the residual vs. fitted value plot gives reason to believe that there could be a minor issue of heteroscedasticity as the variance increases towards zero. Generally, the residual plots give reason to believe that we have unbiased estimators for predicting normal returns. Potential heteroscedasticity could however be a minor issue.

Figure A1.2: Residual Plots for BW Offshore Ltd Listed at Oslo Børs Stock Exchange in Norway.

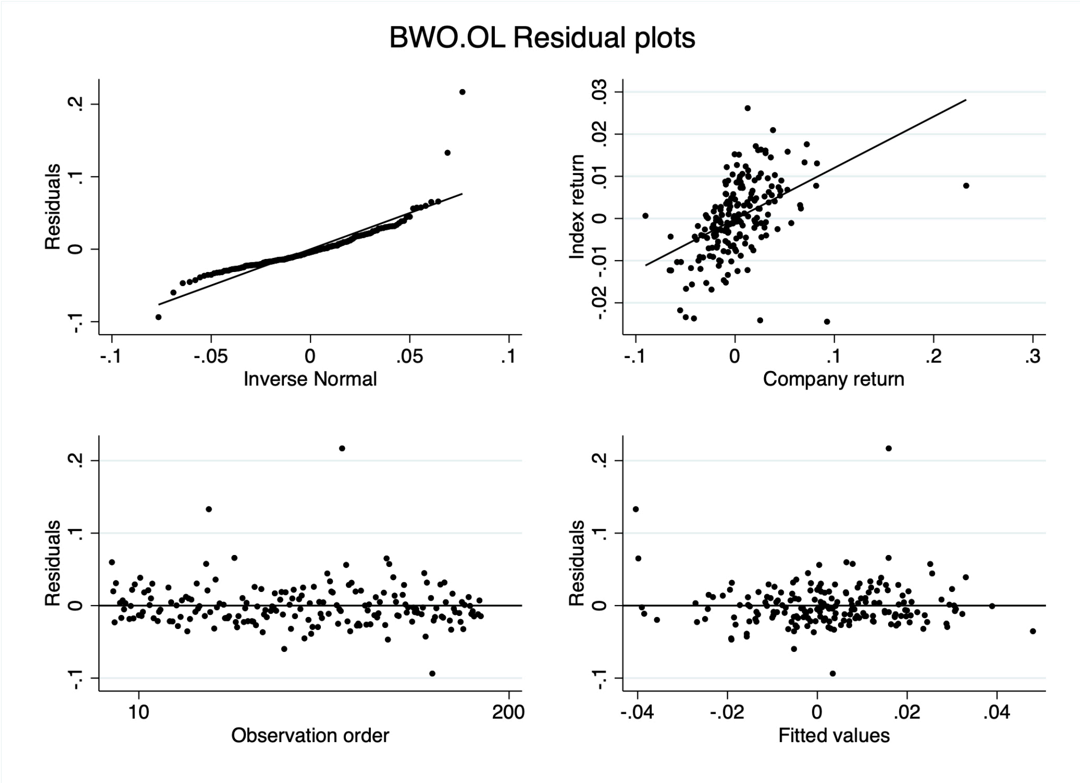


Figure A1.2 shows four different residual plots for BW Offshore Ltd regressed on the Value Weighted Index (OMXO20GI), in the estimation window [-200,-10] for event date 31.12.2019. The upper left plot is the normal probability plot (qnorm), which indicates whether the residuals are normally distributed or not. To the upper right is a linearity plot where we test whether the linearity assumption between firm returns and index returns hold. To the lower left is a residuals vs. observations order plot, which indicates whether the residuals are uncorrelated or not. To the lower right is the residual vs. fitted value plot, which indicates whether the residuals have constant variance (homoscedasticity) or not (heteroscedasticity).

Deviations from the trend line are minor, and the normality assumption is accepted. There are a few observed outliers in the linearity plot, but observations do not deviate much from linearity. No clear relationship between residuals is identified in the residual vs. observation order plot, and we find the correlation assumption reasonable. Although there are some outliers in the residual vs. fitted values plot, the plot gives reason to believe that the constant variance assumption can be accepted. Generally, the residual plots give reason to believe that we have unbiased estimators for predicting normal returns, as there are no major problems for the assumptions.

Figure A1.3: Residual Plots for EAC Invest AS Listed at the Nasdaq Copenhagen Stock Exchange in Denmark.

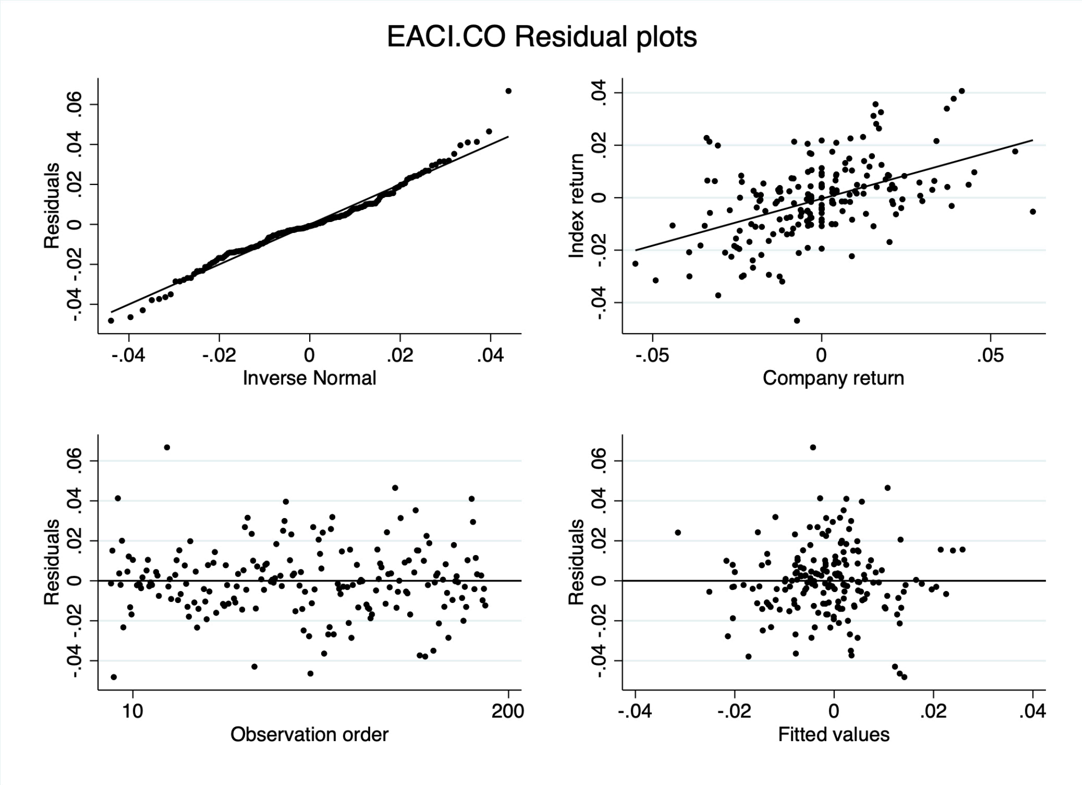


Figure A1.3 shows four different residual plots for EAC Invest AS regressed on the Value Weighted Index (OMXC20), in the estimation window [-200,-10] for event date 31.12.2011. the upper left plot is the normal probability plot (qnorm), which indicates whether the residuals are normally distributed or not. To the upper right we test whether the linearity assumption between firm returns and index returns hold. To the lower left is a residuals vs. observations order plot, which indicates whether the residuals are uncorrelated or not. To the lower right is the residual vs. fitted value plot, which indicates whether the residuals have constant variance (homoscedasticity) or not (heteroscedasticity).

Deviations from the trend line are minor, and the normality assumption is accepted. There are a few observed outliers in the linearity plot, observations do not to deviate too much from linearity. No clear relationship between residuals are identified in the residual vs. observation order plot, and the correlation assumption is accepted. Although there are some outliers, the residual vs. fitted values plot gives reason to believe that the constant variance assumption can be accepted. Generally, the residual plots give reason to believe that we have unbiased estimators for predicting normal returns, as there are no major problems present.

Figure A1.4: Residual Plots for Caverion Group Listed at the Nasdaq Helsinki Stock Exchange in Finland.

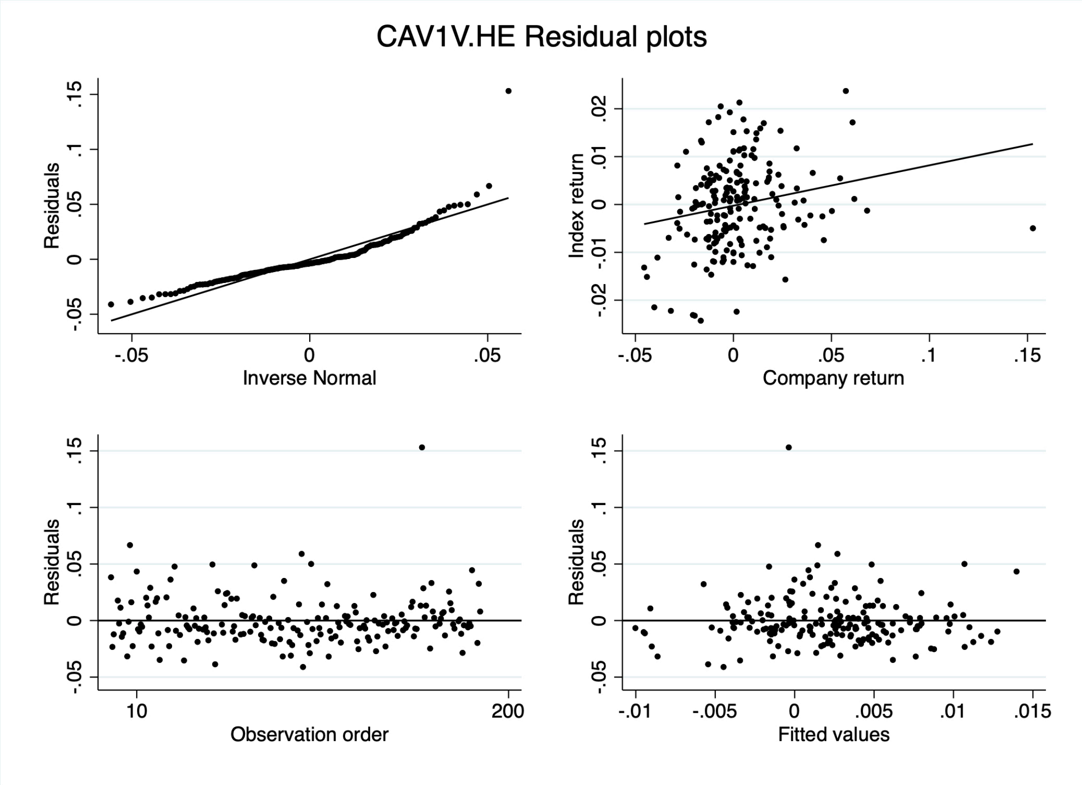


Figure A1.4 shows four different residual plots for Caverion Group regressed on the Value Weighted Index (OMXH25), in the estimation window [-200,-10] for event date 31.12.2019. Upper left figure is the normal probability plot (qnorm), which indicates whether the residuals are normally distributed or not. To the upper right we test whether the linearity assumption between firm returns and index returns hold. To the lower left is a residuals vs. observations order plot, which indicates whether the residuals are uncorrelated or not. To the lower right is the residual vs. fitted value plot, which indicates whether the residuals have constant variance (homoscedasticity) or not (heteroscedasticity).

Deviations from the trend line in the normality plot are minor, and the normality assumption can be accepted. Observations in the linearity plot indicate a linear relationship between company returns and index returns, and the linearity assumption seems reasonable. As no clear relationship between residuals is identified in the residual vs. observation order plot, we find the correlation assumption to hold. Apart from the observed outliers, the residuals vs. fitted values plot gives reason to believe that the constant variance assumption can be accepted. Generally, the residual plots give reason to believe that we have unbiased estimators for predicting normal returns, as there are no major problems present.

Table A1.1: Summary Statistics for Market Model Residuals.

Ticker	Event date	Breuch Pagan	Durbin Watson	Normality
CARLb.CO*	02/01/2012	0.0335	1.640	OK
AAK.ST	02/01/2020	0.9989	1.343	OK
AKAS.OL	02/01/2013	0.5175	2.007	OK
COLOb.CO	30/09/2015	0.2000	1.635	OK
SECTb.ST	30/04/2019	0.3360	1.635	OK
BOL.ST	02/01/2012	0.0885	1.837	OK
CGCBV.HE*	04/01/2016	0.0016	1.826	OK
HUH1V.HE	02/01/2020	0.6901	1.886	OK
DNO.OL	02/01/2012	0.7735	1.711	OK-
BO.CO	03/06/2019	0.6566	1.887	OK
DSV.CO	02/01/2015	0.3253	1.661	OK-
MAERSKb.CO	04/01/2016	0.9060	1.636	OK
KNEBV.HE	02/01/2012	0.6813	1.977	OK
EKTAb.ST	30/04/2013	0.5423	1.437	OK
NCCb.ST	02/01/2017	0.3632	1.516	OK
BERGb.ST	03/04/2018	0.6547	1.587	OK
DNB.OL*	02/01/2018	0.6917	1.679	OK
GJES.OL	02/01/2017	0.0924	1.709	OK
UPONOR.HE*	02/01/2018	0.2953	1.669	OK
REC.OL	02/01/2017	0.9248	1.811	OK
KCRA.HE*	02/01/2019	0.8978	1.788	OK

Table A1.1 presents test results for residuals to check OLS assumptions for the market model. Residuals are tested for 20 different estimation windows, where firm returns are regressed on index returns. All estimation windows range from [-200,-10] relatively to the event date presented. Breuch Pagan: p-values to check for for homoscedasticity. Null hypothesis is constant variance (homoscedasticity). Durbin Watson: Test statistic to check for independece. n = 190 and one independent variable (market index) gives lower critical value (dL) of 1.758 at the 5 percent level. A Durbin Watson statistic below 1.758 means there is evidence of positive serial correlation, and if 4 minus the Durbin-Watson statistic is below 1.758, there is evidence of negative serial correlation. Normality: Check for normality assumption. The check is done by examining histograms of the residuals. Presenting all histograms in the appendix should be of minimal use to the reader, but they are available upon request. OK means that the normality assumption is accepted. OK- means that histograms are slightly skewed to the right. KCRA.HE*: 1 outlier removed. Original Breuch Pagan value 0.0043. UPONOR.HE*: 2 outliers removed. Original Brech Pagan value 0.0000. DNB.OL*: 4 outliers removed. Original Breuch Pagan value 0.0001. CGCBV.HE*: 2 outliers removed. Original Breuch Pagan value 0.0016. CARLb.CO*: 1 outlier removed. Original Breuch Pagan value 0.0335.

Table A1.1 is used to test whether OLS assumptions hold for the market model.There is evidence of heteroscedasticity at the 5 percent level for five estimation windows, but there is no evidence when a few outliers are removed. Thus, the homoscedasticity assumption seems reasonable. About half of estimation windows have some evidence of positive serial correlations, but problems seem minor. The normality assumption seems to hold.

Table A1.2: Summary of Regression Results for Defining the Market Models.

Norway	Ticker	$\beta 1$	cons	Adj.R^2	N
Main-index	OSEBX	1.174***	-0.000	0.188	45102
Value weighted	OMXO20GI	1.100***	-0.000***	0.183	45102
Sweden					
All shares	OMXSGI	0.987***	-0.000***	0.267	105687
Mid-cap	OMXSMCGI	1.072***	-0.000***	0.222	105687
Large-cap	OMXSLCGI	0.952***	0.000***	0.265	105687
Industrials	SX50GI	0.749***	-0.000***	0.242	105687
Value weighted	OMXS30NEXTGI	0.935***	-0.000**	0.251	105687
Finland					
Mid-cap	OMXHMCGI	1.2341***	-0.000**	0.270	42750
Large-cap	OMXHLCGI	1.0206***	-0.000**	0.337	42750
Industrials	HX50GI	0.895***	-0.000**	0.314	42750
Value weighted	OMXH25	1.002***	-0.000	0.345	42750
Denmark					
Mid-cap	OMXCMCGI	0.927***	0.000	0.116	46462
All shares index	OMXCGI	0.932***	-0.000	0.144	46462
Value weighted	OMXC20	0.794***	-0.000	0.128	46462

Table A1.2 holds summary regression statistics for defining market models for Norway, Sweden, Finland, and Denmark. Daily stock returns in the estimation window [-200,-10] for all defined events from 2011 to 2020 are regressed on respective daily index returns. Ticker: Index identifier. $\beta 1$: Regression estimator. cons: Regression constant. Adj.R^2: Measure of regression explanatory power. N: Number of observations examined. OSEBX: The Norwegian main index. OMXO20GI: The Norwegian value weighted index consisting of the 20 most traded securities. OMXSGI: The Swedish all-shares index. OMXSMCGI: Swedish mid-cap index. OMXSLCGI: The Swedish large-cap index. SX50GI: The Swedish industrials industry index. OMXS30NEXTGI: The Swedish value weighted index consisting of the 30 most traded securities in Sweden. OMXHMCGI: The Finish mid-cap index. OMXHLCGI: The Finish large-cap index. HX50GI: The Finish industrials industry index. OMXH25: The Finish value weighted index consisting of the 25 most traded shares in Finland. OMXCMCGI: The Danish mid-cap index. OMXCGI: the Danish all shares index. OMXC20: The Danish value weighted index consisting of the 20 most traded securities in Denmark. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.

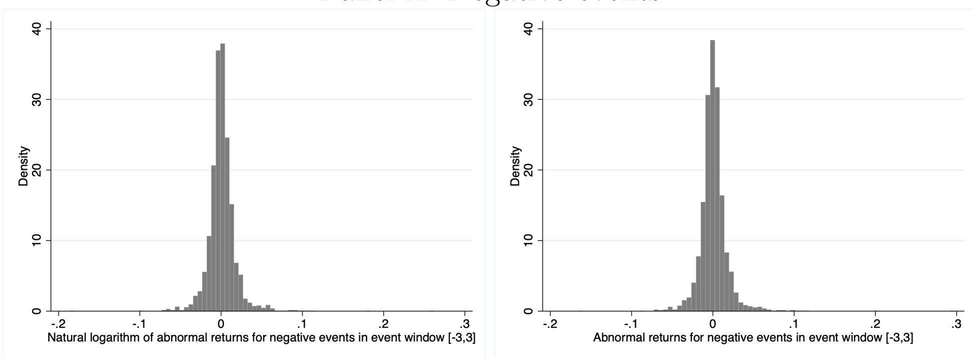
Table A1.2 presents test results for stock returns regressed on indices to determine best fit for market models. For Sweden, the large-cap index and value-weighted index have highest explanatory power with R^2 respectively of 26.5% and 25.1%. Similar for Finland, with R^2 of 33.7% for the large-cap index and 34.5% for the value weighted index. For Denmark, the all shares index and the value weighted index has highest explanatory power with R^2 of respectively 14.4% and 12.8%. For Norway, the main index and the value weighted index have R^2 of 18.8% and 18.3%.

A2 Appendix B

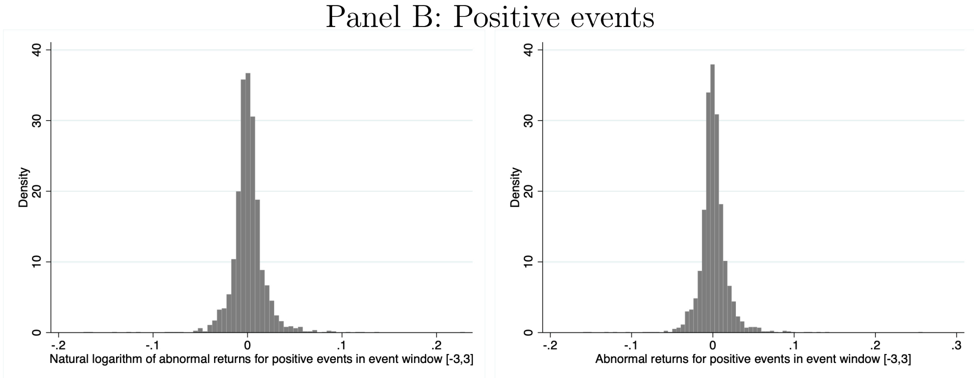
A2.1 Normality for Event Window Observations

Figure A2.1: Summary of Simple and the Natural Logarithm of Abnormal Returns for all Events.

Panel A: Negative events



Panel B: Positive events



Panel C: Neutral events

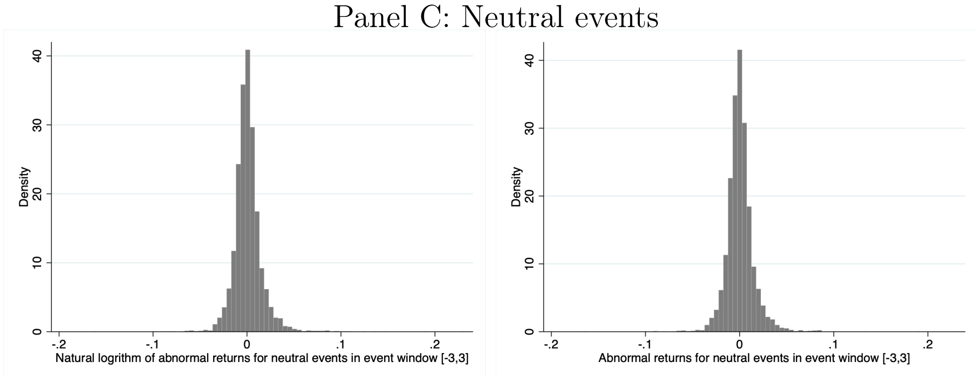


Figure A2.1 shows the distribution of simple and log abnormal returns in the event window [-3, 3] associated with an ESG score change. Panel A shows the distribution for events categorized as negative. Negative events are identified as ESG score y/y change of -2.77% or more. Panel B shows the distribution for events categorized as positive. Positive events are identified as ESG score year-on-year changes of more than 10.61%. Panel C shows the distribution for events categorized as neutral. Neutral events are identified as y/y change lower than 10.61% and higher than -2.77%.

A2.2 Market Capitalisation Guidelines

Table A2.1: Market Capitalisation Guidelines.

Panel A: Nasdaq Nordic		
	Lower Bound (EUR)	Upper Bound (EUR)
Small Cap	-	150 000 000
Mid Cap	150 000 000	1 000 000 000
Large Cap	1 000 000 000	-
Panel B: Oslo Børs		
	Lower Bound (NOK)	Upper Bound (NOK)
Small Cap	-	1 000 000 000
Mid Cap	1 000 000 000	15 000 000 000
Large Cap	15 000 000 000	-

Table A2.1 summarises the market capitalisation guidelines applied. For the Swedish, Finish, and Danish market, they are defined by the Nasdaq Nordic guidelines defined by the latest yearly Market Cap Segment Review. Firms listed at Oslo Børs follow the guidelines provided by Oslo Børs. Market capitalization were retrieved in USD, and converted to following exchange rates based on rates as of 5th October 2020. USD/EUR = 0.85, USD/NOK = 9.28.

A2.3 Clustered Event Dates

Table A2.3: Summary Statistics for Observed ESG Score Changes 2011-2020

Event date	Negative	Neutral	Positive	Event date	Negative	Neutral	Positive
03/01-11	2	2	0	31/12-15	29	53	35
29/04-11	0	0	1	29/04-16	0	1	0
30/05-11	0	1	0	30/05-16	0	1	0
30/08-11	1	0	0	30/08-16	0	1	0
29/09-11	1	0	0	29/09-16	0	1	0
30/11-11	1	0	0	30/11-16	1	0	1
31/12-11	35	43	31	31/12-16	36	64	31
29/04-12	0	0	1	31/03-17	1	0	0
30/05-12	0	0	1	29/04-17	1	2	0
30/08-12	0	0	1	30/05-17	0	1	0
29/09-12	0	1	0	30/08-17	0	1	0
30/11-12	1	0	1	29/09-17	1	0	0
31/12-12	39	45	26	01/11-17	0	1	0
29/04-13	0	0	1	30/11-17	0	2	0
30/05-13	0	1	0	31/12-17	33	75	23
30/08-13	0	1	0	01/03-18	1	0	0
29/09-13	0	1	0	30/03-18	0	0	1
01/11-13	0	1	0	29/04-18	0	2	1
30/11-13	1	0	0	30/05-18	1	0	0
30/12-13	0	1	0	30/08-18	0	2	0
31/12-13	32	63	14	29/09-18	0	2	0
30/04-14	1	0	0	01/11-18	0	1	0
30/05-14	0	1	0	30/11-18	0	0	2
30/08-14	0	1	0	31/12-18	13	89	41
29/09-14	0	1	0	01/03-19	0	1	0
01/11-14	0	1	0	30/03-19	0	0	1
30/11-14	0	1	1	29/04-19	0	1	2
31/12-14	37	51	24	30/05-19	0	0	1
30/04-15	0	1	0	30/08-19	1	1	0
30/05-15	1	0	0	29/09-19	1	1	1
30/08-15	0	0	1	30/11-19	1	0	1
29/09-15	0	0	1	31/12-19	49	116	70
02/11-15	0	1	0	25/02-20	1	0	0
30/11-15	0	2	0	29/02-20	0	0	1
				30/03-20	0	1	0

Table A2.3 summarises every date an ESG Score change, meaning an announcement for a firm which had a pre-existing ESG Score, was observed in the time period 2011 to 2020 for our sample of Nordic firms. Positive events are defined as ESG score year-on-year change more than 10.61%. Negative events are defined as y/y change of less than -2.77%. Neutral events are defined as y/y change less than 10.61% and more than -2.77%.

A2.4 Industry Segmentation Guide

Table A2.5: Industry Segmentation by Thomson Reuters Standards.

Energy	*Real estate*
Energy - Fossil fuels	Real estate
Renewble energy	Collective investments
Uranium	Investment holding companies
Materials	*Technology*
Chemicals	Technology equipment
Mineral resources	Software and IT services
Applied resources	Telecommunication services
Industrials	*Healthcare*
Industrial goods	Healthcare services and equipment
Industrial and commercial services	Pharmaceuticals and medical research
Consumer goods conglomerates	
Transportation	
Consumer cyclicals	*Academic and educational services*
Automobiles and auto parts	
Cyclical consumer products	
Cyclical consumer services	
Retailers	
Consumer non cyclicals	*Government activity*
Food and beverages	
Personal and household products and services	
Food and drug retailing	
Financials	*Institutions associations and organizations*
Banking and investment services	
Fintech and infrastructure	
Insurance	
Utilities	

Table A2.4 displays the industry segmentation guide by Thomson Reuters.

A2.5 Original BMP Test and Rank Test Results

Table A2.7: Summary Statistics for Original BMP Test and Rank Test.

Panel: A - Positive events

Event window	N	ACAR	BMP	Adj.BMP	C&Z	GRANKT
[-5,5]	315	0.0114	2.9516***	2.1461**	1.6673*	1.5004
[-3,3]	315	0.0067	2.4325**	1.7692*	1.3044	2.4176**
[-1,1]	311	0.0040	2.0539**	1.4938	1.2889	1.4857
[-5,0]	316	0.0105	3.4295***	2.4943**	2.1191**	1.4153
[-3,0]	316	0.0077	3.5679***	2.5950**	2.0816**	2.9044**
[-1,0]	316	0.0057	3.2751***	2.3821**	1.8135*	3.1446**
[0,1]	310	0.0031	1.6732*	1.2170	1.2842	1.4967
[0,3]	314	0.0037	1.2189	0.8865	0.6933	0.7323
[0,5]	315	0.0052	1.4401	1.0474	0.9737	1.2798
[0,10]	315	0.0046	1.4012	1.0191	0.5945	1.5746

Panel: B - Negative events

Event window	N	ACAR	BMP	Adj.BMP	C&Z	GRANKT
[-5,5]	321	0.0115	2.6281***	1.8911*	1.6900*	2.9962***
[-3,3]	319	0.0064	1.8124*	1.3041	1.0328	1.3998
[-1,1]	321	0.0060	2.5606**	1.8425*	1.1239	2.5557**
[-5,0]	322	0.0090	2.5909**	1.8643*	1.9473*	3.2648***
[-3,0]	322	0.0059	2.2303**	1.6049	1.5226	2.8375***
[-1,0]	322	0.0053	2.9153***	2.0978**	1.5450	1.6847*
[0,1]	320	0.0030	0.9160	0.6591	0.1557	0.9855
[0,3]	318	0.0026	0.4762	0.3426	-0.1409	0.5441
[0,5]	321	0.0047	1.1427	0.8222	0.4871	1.9840**
[0,10]	321	0.0044	0.4890	0.3519	-0.4798	0.8694

Panel: C - Neutral events

Event window	N	ACAR	BMP	Adj.BMP	C&Z	GRANKT
[-5,5]	639	0.0065	1.3258	0.9130	0.3773	1.4821
[-3,3]	637	0.0052	1.6676*	1.1484	0.0330	-0.0494
[-1,1]	637	0.0032	1.6896*	1.1635	-0.1024	0.0173
[-5,0]	640	0.0043	1.4764	1.0167	1.1268	1.1065
[-3,0]	640	0.0050	2.5403**	1.7494*	1.2375	1.6939*
[-1,0]	640	0.0037	2.7925***	1.9231*	0.7533	1.1376
[0,1]	635	0.0012	-0.3368	-0.2319	-0.9553	-0.0750
[0,3]	635	0.0017	-0.3018	-0.2078	-1.3705	-1.9289*
[0,5]	637	0.0038	0.2240	0.1542	-0.7445	-0.4780
[0,10]	637	0.0022	-0.2559	-0.1762	-1.1367	-0.9883

Table A2.7 summarises test results for positive, negative, and neutral events. Average cumulative abnormal returns (ACAR) are associated with the defined event window for the respective event day. Positive events are defined as ESG Score year-on-year $>$ 10.61%. Negative events are defined as y/y change $<$ -2.77%. Neutral events are defined as y/y change less than 10.61% and more than -2.77%. H_0 of no abnormal returns is tested with the original BMP test (BMP)(Boehmer et al., 1991), adjusted BMP test (Adj.BMP)(Kolari and Pynnönen, 2010), Cumulative Rank N-test (C&Z)(Corrado and Zivney, 1992) and the Generalized Rank T-test (GRANK)(Kolari and Pynnonen, 2011). N: Number of events. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.

A3 Appendix C

A3.1 Probit Regression Correlation

Table A3.1: Correlation Matrix of Independent Variables for Probit and Multionomial Logit Regressions.

Panel: A - Positive events

	Ebitda	*ROA*	*Beta*	*ln.asset*	*ln.mcap*	*ESG.d*	*Fossil.d*	*Bank.d*	*Time.d*
Ebitda	1.000								
ROA	0.265	1.000							
Beta	0.048	0.033	1.000						
ln.asset	0.331	-0.039	0.116	1.000					
ln.mcap	0.258	0.261	0.140	0.705	1.000				
ESG.d	0.061	0.098	0.138	0.245	0.336	1.000			
Fossil.d	0.195	-0.116	0.386	0.007	-0.109	0.015	1.000		
Bank.d	0.175	-0.101	0.001	0.397	0.070	-0.012	-0.099	1.000	
Time.d	0.023	0.021	-0.071	-0.120	-0.135	-0.035	0.006	0.055	1.000

Panel: B - Negative events

	Ebitda	*ROA*	*Beta*	*ln.asset*	*ln.mcap*	*ESG.d*	*Fossil.d*	*Bank.d*	*Time.d*
Ebitda	1.000								
ROA	0.156	1.000							
Beta	-0.007	-0.143	1.000						
ln.asset	0.151	-0.155	0.075	1.000					
ln.mcap	0.068	0.236	0.058	0.752	1.000				
ESG.d	0.069	0.145	0.044	0.305	0.410	1.000			
Fossil.d	0.028	-0.208	0.421	-0.058	-0.123	-0.104	1.000		
Bank.d	0.035	-0.141	-0.072	0.558	0.222	0.005	-0.120	1.000	
Time.d	-0.070	0.092	-0.078	-0.168	-0.095	0.046	-0.043	0.015	1.000

Table A3.1 show the correlation matrix between the independent variables used both for the probit and the multinominal regression. Ebitda: EBITDA margin (Percentage measure calculated by EBITDA/total sales for firm i). ROA: Return on assets (Percentage metric calculated by income after taxes for fiscal year (FY) divided by average total assets of FY). ln.asset: Natural logarithm of average total assets in FY. ln.mcap: Natural logarithm of market capitalization at the end of FY (Calculated by last share price of FY*total shares outstanding). ESG.d: Dummy variable which equals 1 if ESG scores >50 for previous year, and 0 otherwise. Fossil.d: Dummy variable equal 1 if company i operates in the fossil fuel sector. Bank.id: Dummy variable equal 1 if company i operates in the banking sector. Time.d: Time dummy that equals 1 if the observation was in 2017 or later. Correlation is calculated by the Pearson correlation coefficient.

A3.2 Probit Regression Marginal Effects

Table A3.3: Marginal Effects From Probit Regression for Positive and Negative Events.

	Dependent variable (1 = Positive CAR, 0 = Negative CAR)			
Panel A: Positive events				
	Event window [-1,0]		Event window [-3,0]	
	Model(1)	Model(2)	Model(3)	Model(4)
Beta	-0.0485	-0.0264	0.0068	0.0437
ROA	-0.1851	0.0343	0.0064	0.2123
Ebitda	-0.1127	-0.0799	0.0271	0.0529
ln.asset	-0.0387*	-	-0.0424*	-
ln.mcap	-	-0.0725***	-	-0.0983***
ESG.d	-0.0154	0.0118	-0.0174	0.0303
Fossil.d	0.1269	0.0831	0.1616	0.0950
Bank.d	0.1134	0.0498	-0.0248	-0.0901
Time.d	0.1475***	0.1370**	0.1725***	0.1577***
Panel B: Negative events				
	Event window [-1,0]		Event window [-5,0]	
	Model(1)	Model(2)	Model(3)	Model(4)
Beta	-0.0286	-0.0204	-0.1336*	-0.1398**
ROA	-1.0776***	-0.7107**	-0.5113*	-0.2624
Ebitda	0.1662	0.1449	-0.0437	-0.0463
ln.asset	-0.0552***	-	-0.0569***	-
ln.mcap	-	-0.0695***	-	-0.0414*
ESG.d	-0.0868	-0.0776	-0.0732	-0.0940
Fossil.d	-0.0432	-0.0461	0.0471	0.0484
Bank.d	0.1731	0.0906	0.1132	-0.0128
Time.d	0.1628***	0.1713***	0.0084	0.0255

Table A3.3 summarises marginal effects from probit regression for event window [-1,0] and [-3,0] for positive events and event window [-1,0] and [-5,0] for negative events. Positive events are identified with an ESG Score year-on-year change of more than 10.61%. Negative events are identified with y/y change of -2.77%. The dependent variable is a binary variable with 2 possible outcomes where 1 = Positive cumulative abnormal return (CAR), and 0 = Negative CAR. Marginal effects are estimated by: $\frac{\partial Y}{\partial X_i} = \beta_i \phi(\beta_0 + \beta_1 Beta + \beta_2 Ebitda + \beta_3 ROA + \beta_4 ln.asset + \beta_5 ln.mcap + \beta_6 ESG.d + \beta_7 Fossil.d + \beta_8 Bank.id)$. Beta: Stock volatility(stock return variation in relation to market return variation) in estimation window for firm i. Ebitda: EBITDA margin (Percentage measure calculated by EBITDA/total sales for firm i). ROA: Return on assets (Calculated as income after taxes for fiscal year (FY) divided by average total assets of FY). ln.asset: Natural logarithm of average total assets in FY. ln.mcap: Natural logarithm of market capitalization at the end of FY (Calculated by last share price of FY*total shares outstanding). ESG.d: Dummy variable which equals to 1 if ESG scores >50 for previous year, and 0 otherwise. Fossil.d: Dummy variable equal to 1 if company i operates in the fossil fuel sector. Bank.id: Dummy variable equal to 1 if company i operates in the banking sector. Time.d: Time dummy that equals 1 if the observation was in 2017 or later. Model (1): Variable *ln.mcap* excluded. Model (2): Variable *ln.asset* is excluded. Model (3): Variable *ln.mcap* is excluded. Model (4): Variable *ln.asset* is excluded. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.

A4 Appendix D

A4.1 Multinominal Regression

In order to investigate whether specific firm characteristics are more likely to be affiliated with positive or negative events, logit regression is applied. Logit regression is a maximum likelihood technique, which uses explanatory variables to maximise the probability of observing a discrete outcome.

A logit model that allows for multiple outcomes in the dependent variable, which is the case here with positive, negative and neutral events, is called a multinominal logit model (Borooah, 2002). Following outcomes in the dependent variable are defined: 0 = Neutral event, 1 = Positive event, and 2 = Negative event. Note that the outcomes are nominal, meaning they are not ordered[33].

The multinominal logit regression does assume independence of irrelevance alternatives (IIA)[34]. This assumption implies that the odds of a given outcome for the dependent variable, is not affected of the addition or removal of a possible outcome outcome (Borooah, 2002). Intuitively, it is reasonable to believe that the likelihood of a specific event is unaffected by removing either the positive, neutral or negative outcome. Furthermore, our three groups surrounds the whole data set, meaning an added outcome would not affect the outcome for the dependent variable either. The IIA assumption is therefore assumed to hold.

Results are presented in Table A4.1. Results indicate that the likelihood of a negative event compared to a neutral event has a negative relationship with $ESG.d$ and $Time.d$. Additionally, results indicate a positive relationship for negative events with $Bank.d$ and $Fossil.d$. The likelihood of a positive event has a negative relationship with the $ESG.d$, and results indicate a positive relationship with ROA. There is also some weak evidence of a negative relationship between $ln.mcap$ and positive events.

[33] Ordered would have meant outcome 1 is better than 0, and 2 is better than 1.

[34] Other assumptions are independent observations, mutually exclusive outcomes and no multicollinearity for explanatory variables. The first are assumed to hold while the latter two holds as an event can only be classified as either positive, neutral or negative, and no multicollinearity are ensured by choice of explanatory variables.

Table A4.1: Multinominal Logit Regression Analysis.

Firm characteristics associated with positive and negative events
(Parameter estimates)

	0 = Neutral event 1 = Negative event, 2 = Positive event			
Panel A: Negative events				
	Model (1)	Model (2)	Model (3)	Model (4)
Ebitda	-0.1668	-0.1643	-	-
ROA	-	-	0.5737	1.0261
ln.asset	-0.0655	-	-0.0787	-
ln.mcap	-	-0.0683	-	-0.1027*
ESG.d	-0.4734***	-0.4669***	-0.4985***	-0.4626***
Fossil.d	0.4967**	0.4580*	0.4985*	0.4751*
Bank.d	0.7137**	0.5509**	0.7284**	0.5788**
Time.d	-0.5801***	-0.5672***	-0.5778***	-0.5669***
cons	1.3219	1.3695	1.5539	1.2570
Panel B: Positive events				
Ebitda	0.2405	0.2195	-	-
ROA	-	-	0.4528*	1.9316**
ln.asset	-0.1161*	-	-0.0976	-
ln.mcap	-	-0.0874	-	-0.1190*
ESG.d	-1.9066***	-1.9323***	-1.9453***	-1.9243***
Fossil.d	0.2296	0.1719	0.3707	0.3370
Bank.d	0.4316	0.1640	0.5182	0.3370
Time.d	0.1004	-0.0685	-0.0978	-0.0851
cons	2.7976	2.1682	2.3636*	2.7759**
N	1250	1250	1250	1250
Pseudo R^2	0.0863	0.0857	0.0854	0.0858

Table A4.1 summarises the results for the multinomial regression analysis. The dependent variable is a multinominal variable with 3 possible outcomes where 0 = Neutral event, 1 = Negative event and 2 = Positive event. Neutral events are defined as base for the regression. Results do for this reason interpreted relative to neutral events. Positive events are identified with an ESG score year-on-year change of more than 10.61%. Negative events are identified with y/y change of -2.77%. Neutral events are identified as y/y change lower than 10.61% and higher than -2.77%. *Ebitda*: EBITDA margin (Percentage measure calculated by EBITDA/total sales for firm *i*). *ROA*: total return on assets (Percentage metric calculated by income after taxes for fiscal year divided by average total assets of the fiscal year (FY)). *ln.asset*: Natural logarithm of average total assets of the FY. *ln.mcap*: Natural logarithm of market capitalization at the end of FY (Calculated by: Last share price of FY*total shares outstanding). *ESG.d*: Dummy variable which equals to 1 for ESG scores >50 for previous FY and 0 otherwise. *Fossil.d*: Dummy variable equal to 1 if company *i* operates in the fossil fuel sector. *Bank.id*: Dummy variable equal to 1 if company *i* operates in the banking sector. Time.d: Time dummy that equals 1 if the observation was in 2017 or later. cons: Regression constant. N: Total number of observations. Pseudo R^2: Mc Faddens R^2. Model (1): Multinomial regression without the independent variable *ln.mcap*. Model (2): Multinomial regression without the independent variable *ln.asset*. Model (3): Variables *ROA* and *ln.mcap* are excluded. Model (4): Variables *ESG.d* and *ln.asset* are excluded. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.