Does Mandatory Sustainability Reporting Decrease ESG-rating Disagreement?

A Difference in Differences study on the EU Non-Financial Reporting Directive

HUGO CHRISTENSSON

ELIAS HORN MAURIN

Bachelor Thesis

Stockholm School of Economics

2022

Abstract:

ESG-ratings are explicitly incorporated in investment decisions by investors with $121 trillion AUM. Ratings are based on diverse sustainability data, which is collected and synthesized by raters[1] with different methodologies. The scores for identical stocks across raters have an observed correlation of 0.54 on average, hence denoted ESG-rating disagreement. This paper asks whether the correlation has improved over time and examines the causal effect of introducing mandatory sustainability reporting on rating disagreement. We apply a difference in differences approach, exploiting that Sweden is subjected to the EU Directive on Non-Financial Reporting (2014/95/EU) while Switzerland is not. We find that a mandatory sustainability reporting requirement did not significantly predict a change in ESG disagreement in general, with the exception for the social score correlation that had an average treatment effect of -0.23. This result should be interpreted conservatively since the empirical setting deviates from the ideal one, with a flawed comparability assumption. The insignificance of the tests could suggest that a broad sustainability reporting requirement has a negligible effect on ESG-score correlation at the country level, with implications for policymakers who aspire to decrease the ESG-rating disagreement.

Keywords:

Corporate Social Responsibility, ESG Rating Disagreement, ESG Rating Agencies, Regulation, Disclosure

Authors:

Hugo Christensson (24715)
Elias Horn Maurin (24781)

Tutor:

Marieke Bos, Economist and Deputy Director at the Swedish House of Finance at the Stockholm School of Economics

Examiner:

Adrien d'Avernas, Assistant Professor at the Swedish House of Finance at the Stockholm School of Economics

Bachelor thesis
Bachelor Program in Business and Economics
Stockholm School of Economics

[1] For simplicity, we apply the word raters as an equivalent for rating agency and rating provider.

Does Mandatory Sustainability Reporting Decrease ESG-rating Disagreement?

A Difference in Differences study on the EU Non-Financial Reporting Directive

Hugo Christensson Elias Horn Maurin*[2]

Final draft: May 2022
First draft: February 2022

Abstract

ESG-ratings are explicitly incorporated in investment decisions by investors with $121 trillion AUM. Ratings are based on diverse sustainability data, which is collected and synthesized by raters[3] with different methodologies. The scores for identical stocks across raters have an observed correlation of 0.54 on average, hence denoted ESG-rating disagreement. This paper asks whether the correlation has improved over time and examines the causal effect of introducing mandatory sustainability reporting on rating disagreement. We apply a difference in differences approach, exploiting that Sweden is subjected to the EU Directive on Non-Financial Reporting (2014/95/EU) while Switzerland is not. We find that a mandatory sustainability reporting requirement did not significantly predict a change in ESG disagreement in general, with the exception for the social score correlation that had an average treatment effect of -0.23. This result should be interpreted conservatively since the empirical setting deviates from the ideal one, with a flawed comparability assumption. The insignificance of the tests could suggest that a broad sustainability reporting requirement has a negligible effect on ESG-score correlation at the country level, with implications for policymakers who aspire to decrease the ESG-rating disagreement.

Keywords: Corporate Social Responsibility, ESG Rating Disagreement, ESG Rating Agencies, Regulation, Disclosure

JEL Classification Codes: M14, G14, G24, G28

[2] Christensson is a Bachelor Student at the Stockholm School of Economics, e-mail: 24715@student.hhs.se, Horn Maurin is a Bachelor Student at the Stockholm School of Economics, e-mail: 24781@student.hhs.se. We thank Marieke Bos, the deputy director of the Swedish House of Finance at Stockholm School of Economics for her thoughtful and valuable input. We also thank our fellow students, Haimov and Holmberg, who have had the patience to read and review our paper. We thank Mercer Sweden for providing access to the Bloomberg terminal and the Swedish House of Finance for providing access to the Thomson Reuters Eikon database.

[3] For simplicity, we apply the word raters as an equivalent for rating agency and rating provider.

1. Introduction

"Today, investors increasingly want to understand the climate risks of the companies whose stock they own or might buy. Large and small investors, representing literally tens of trillions of dollars, are looking for this information to determine whether to invest, sell, or make a voting decision one way or another.

Investors are looking for consistent, comparable, and decision-useful disclosures so they can put their money in companies that fit their needs."

- SEC Chairman Gary Gensler, July 28, 2021

Motivation and Question

ESG-ratings are explicitly incorporated in investment decisions by investors with $121 trillion AUM (PRI 2021). Furthermore, ESG-score consideration is mandatory within due diligence processes for six of the seven Swedish Pension AP funds (Swedish Ministry of Finance 2021). Investors have criticized ESG raters for their high inaccuracies and lack of coherence (Wong and Petroy 2020). The inaccuracy is confirmed in prior research which found an average correlation of 0.54 (Berg, Koelbel and Rigobon 2022) and 0.46 (Gibson, Kreuger and Schmidt 2020) between the major raters, which has unsettling implications for whether the right investment decisions are being made. A natural question that follows is whether rating disagreement has decreased over recent years due to increased availability and comparability of ESG-related information after the introduction of the mandatory non-financial reporting directive in the EU (2014/95/EU). That is the question we address in this paper.

Conceptual Framework

To answer whether ESG-rating disagreement has increased- or decreased over time, we need to understand the variables which influence the difference between ESG-ratings for identical stocks across raters. ESG-raters act as third-party assessors of a corporation's total-, environmental-, social- and governance sustainability performance. The performance is measured by multiple underlying indicators of attributes that the rater decides the relative importance of when approximating a score. These indicators are based on several sources, where the annual (sustainability) report is the primary. This score is either relative to the industry or the entire investing universe.

The Non-Financial Reporting Directive (NFRD) has the explicit purpose of making large firms' sustainability reporting more available and comparable (Directive 2014/95/EU). When firms are subjected to a mandatory reporting requirement, it should intuitively enable ESG-raters to improve the interpretation of the firms' sustainability performance, and thus be more

coherent and accurate in their measurements. Berg et al. (2019) found that measurement error represents 56 % of ESG-rating disagreement, which validates the importance of improving the accuracy of measurement. We exploit that Sweden must adhere to the EU Directive on Non-Financial Reporting (NFRD), while Switzerland must not.

Challenges to Answering the Question

Providing an empirical answer to the question of ESG-rating disagreement faces two key challenges. First, we need to examine the causal effect of introducing mandatory sustainability reporting as opposed to keeping voluntary sustainability reporting. Second, ESG reporting is a relatively new phenomenon which delimits the available historical sample size.

Challenge 1: Causal Effect of Mandatory Reporting Requirement

To examine the causal effect of new regulation, one would ideally compare the treated group with the counterfactual. In this hypothetical world, the perfect scenario would be to randomly assign mandatory sustainability reporting equally spread in all sectors and regions and then examine the causal effect. However, since the counterfactual does not exist, we need to construct a natural experiment where we can assume that the control group is comparable enough for us to draw conclusions on the incremental impact. The second-best alternative, and our initial thought when designing the research, was to compare the treated group (i.e., Swedish firms subjected to the NFRD) with an almost identical control group (i.e., Swedish firms not subjected to the NFRD) by exploiting the criteria cut-off based on company size. In practice, we would compare the correlation for firms within Sweden just above the cut-off contrary to firms just below the cut-off before and after the policy change. This was impossible since the low cut-off included firms subjected to NFRD which lacked any ESG rating, hence one could not construct a sufficient control group sample below the cut-off since the firms' in that group was not rated. With this limitation in mind, we figured that Norway had similar country-specific characteristics without an EU-membership, and therefore investigated whether Norway was exempt from the EU Non-Financial Reporting Directive. We found that Norway had introduced a law regarding the disclosure of non-financial information in 2012, thus Norway could not act as control group. As a last resort, we attempt to overcome this challenge by comparing Switzerland and Sweden with the assumption that the countries' have similar macro country-characteristics and a similar investor sentiment with regards to ESG-issues (see Section 3.2.1).

Challenge 2: Data Quality, Availability and Sample Size

As Milton Friedman eloquently once said: "there is one and only one social responsibility of business - to use its resources and engage in activities designed to increase its profits so long as it stays within the rules of the game, which is to say, engages in open and free competition without deception fraud" (the New York Times 1970). This principle has perpetuated business theory while only being substituted by stakeholder theory recently, which was introduced by Edward Freeman in 1984 (Freeman, 2010). Further, the notion of Corporate Social Responsibility (CSR) lagged as a legitimate subject of financial research, with the first CSR-related publication in the Journal of Finance in 2017 (Liang, Renneboog 2017). Hence, the incorporation of other than pure financial risk dimensions in investment decisions is a relatively new phenomenon. This novelty creates implications for the ESG-rater market and its quality of data.

First, the emerging market has no consensus on standard frameworks of methodology or disclosure of methodology. It differs significantly from credit ratings which have straightforward and publicly disclosed rating methodologies, both model- and analyst driven. Both are predominantly based on public financial statements and can be benchmarked over time (S&P Global 2019). There is no ultimate truth in tracking firms' ESG-information as sustainability performance is multifaceted and substantially harder to supervise and analyse.

Second, the ESG-rater market is fragmented, and investors commonly incorporate multiple ESG sources to improve their investment analysis (Wong and Petroy 2020). The fragmentation creates incentives for ESG-raters to differentiate themselves to meet the varying investor preferences within the field of sustainable investment. Consequentially, different methodologies are applied on the same stocks, which lowers the overreaching informational value of ESG-ratings as an indicator of the firm's sustainability performance and confuses investors. Furthermore, as the methodology is a key point of differentiation, changes of rating methodology to improve their services thus become a natural aspect.

This competitional aspect contributes to a fierce restriction on data availability, the low historical interest generates a short historical period, while the improvement of offering in terms of methodology decreases the data quality of the already short period of data. These challenges pose strong limitations in the generalization of our findings.

Results Summary

In general, our results suggest that the introduction of a mandatory sustainability reporting requirement did not have a significant effect on ESG-rating disagreement. Nevertheless, we find a significant negative average treatment effect of -0.23 for the social score correlation on the 5 % significance level. That effect should be interpreted with caution due to the limitations in the empirical setting and assumption of comparability. The tests for the total-, environmental- and governance scores correlation were insignificant with negative point estimates for all coefficients except for governance.

Robustness of Empirical Strategy

To test the robustness of our findings, we constructed two tests to understand whether our findings could be explained by noise in the data or other omitted variables.

First, we ran a test where we assume that the mandatory reporting requirement was introduced in 2014, while keeping 2012 as a baseline year and 2016 as the final year. Thus, we examine the effect of the mandatory reporting requirement before it occurred, which should be zero. All other parameters, such as the treatment and control group, are identical to the original regression. The test was insignificant for all coefficients, which could be partially explained by the decreased sample size when changing the event time from nine to five years.

Second, we reran the original regression with a randomized sample of firms, by randomly assigning a country to each firm while maintaining the original proportions for each hypothetical country. The test was insignificant for the social score correlation while significant for the total score correlation. The significant results were surprising and decreases the robustness of the test.

Mechanism

Three mechanisms could explain our findings. First, the number of reported non-financial indicators on Nordic Compass have decreased from 2014-2017 according to the Swedish Agency for Growth Policy Analysis (2018). If that trend has remained, it could potentially explain why the Non-Financial Reporting Directive in isolation had a negligible effect on ESG-rating disagreement. Second, Bond and Zeng (2021) argue that “Silence is Safest” as a disclosure policy for firms to avoid negative exposure. Thus, the optimal choice for firms exposed to the Non-Financial Reporting Directive might be to avoid disclosing controversial

ESG-information and instead disclose non-material issues instead. Thus, the informational value of reported indicators decreases, and the directive fails to reach its desired effects, which similarly to the first mechanism decreases the causal effect of mandatory reporting on ESG-disagreement. Third, with the proliferation of ESG-raters and investors' interest in their services, the competition has increased in tandem with the need for differentiation. The product is synonymous with their rating methodology, where improvements are a response to investor needs, however with undisclosed changes, it may instead lead to skepticism. Lastly, the revision of rating methodology with simultaneous lack of disclosure and transparency regarding these changes makes ESG-data over time less comparable and increases rating disagreement.

Related Literature

Our paper fits in the financial literature that investigates the effects of corporate social responsibility, prosocial investor sentiment and sustainability on the financial market in terms of portfolio choice, asset pricing (see: Pedersen, Fitzgobbins, Pomorski 2021; Avramov, Cheng, Lioui, Tarelly 2021), stock performance (see: Berg, Koelbel, Pavlova, Rigobon 2021) firm performance, fund investment flows (see: Hartzmark, Sussman 2019) and financial decision making. It also relates to the effects that third-party financial institutions have on the risk profile of individual companies, with a clear similarity between credit ratings and ESG-ratings as a quality assessment of a firm. A selected sample of relevant literature and our contribution is covered in the following paragraphs.

To explain the variation in correlation across raters for the same firms (ESG-rating disagreement) Gibson et al. (2020) have researched the explanatory value of firm-level characteristics, such as: balance sheet, industry, investor transparency, and valuation and their relation to stock returns. However, firm-level characteristics have proven to exhibit a low explanatory value (i.e., 0.047-0.059 in adjusted R^2) for the ESG-rating disagreement. Berg et al. (2019) apply another approach, namely a systematic categorization of divergence into three separate errors, namely scope divergence, measurement divergence and weights divergence. Scope divergence is intuitively the act of deciding what kind of attributes that are in- and out of scope to estimate the ESG-performance of corporations. Measurement divergence relates to the practice of different raters who have the same scope of attributes yet use different indicators to measure these attributes. The weights divergence relates to the relative weights that are assigned to each attribute when aggregating the scores. The paper finds that measurement-,

scope-, and weight errors explain 56 %, 38 % and 6 % of the divergence respectively. The focal point of our paper is to extend the research that deals with the existence, explanation, and implications of rating disagreement across ESG-raters. We contribute to the literature by adding a time dimension that shows the divergence or convergence of the phenomenon, while simultaneously exploring whether a mandatory sustainability reporting requirement has a significant effect on the phenomenon. We consider our contribution to be of value in countries which have yet to introduce such policy mandates, as is the case for the United States.

Liang and Renneboog (2017) studied explanatory factors for ESG-ratings and found that countries' legal origin, i.e., having a common- or civil law system, had the strongest explanatory value for differences in ESG-score globally. Specifically, civil law systems exhibited a higher ESG-score and were also more responsive to CSR shocks. Our study extends this literature through the differentiation and comparison on the effect of legal changes to reporting requirements in two countries which both are subjected to a civil law system, thus increasing the understanding of the role of legal aspects within ESG. Similarly, Eccles and Stroehle (2018) attempts to explain rating disagreement through the lens of social constructionism by analysing ESG-ratings as a function of the raters' values. They investigate ESG-raters' social origins (e.g., founding principles, legal status, purpose) and their fundamental need for differentiating themselves in a competitive and maturing market. They find that the raters' values influence their methodology in terms of perception of sustainability and materiality which affects the marketing of their products. This directly relates to our research as a potential limitation with the number of ESG-raters included in our thesis (see the ideal sample size in Section 2.4)

Lins, Servaes and Tamayo (2017) studied the performance of firms with high social capital, as measured by their ESG ratings, during the financial crisis of 2008-2009. They found that performance was significantly higher (i.e., four to seven percentage points) for firms with higher social capital than low. As an underlying mechanism, they interpreted social capital as a means of exhibiting trust amid a sharp increase in the lack of trust in society. Our thesis relates to their article in two ways. First, the notion of rating disagreement decreases the legitimacy of only using one ESG-rater's scores as an indicator of trust. Second, one potential effect of mandatory reporting regulation could be that the overall average sustainability performance increases, whereas the definition for being recognized as a firm with social capital must be adjusted. This idea is in line with the reasoning of Bénabou and Tirole (2009) and the

signalling effect of engaging and taking ownership of ESG questions. At last, our paper contributes to understanding legal incentives as a factor that predicts ESG engagement on the corporate level.

Paper Structure

The remainder of the paper is structured as follows. Section 2 provides an overview of the empirical setting, describes the data, and presents summary statistics. Section 3 describes our empirical strategy and Section 4 presents the results of the applied empirical strategy. Section 5 describes the underlying mechanisms in our interpretation of the test results. Section 6 concludes our findings.

2. Setting, Data and Summary Statistics

In this section, we describe the setting by elaborating on the reporting requirement, ESG background, rating methodology and sustainable investment strategies. We also report our data sources, sample selection and present descriptive summary statistics.

2.1 The Non-Financial Reporting Directive and Swiss Regulation

The empirical basis for our thesis lies on the Non-Financial Reporting Directive (NFRD) which was introduced within the European Union (EU) in 2014 and entered into effect in 2017 (2014/95/EU). NFRD requires disclosure of non-financial information, such as: environmental issues, CSR, labour rights, human rights, and anti-corruption. NFRD is based on a comply or explain manner. The purpose of NFRD is to increase transparency and comparability of sustainability reports across companies within the EU. NFRD defines a large undertaking (i.e., firm) with above 500 employees as the prescribed minimum level for being subjected to the law. The directive was translated into stricter terms on a national level in Sweden where a company that fulfils at least two of the following three criteria are subjected to the law:

1. On average 250 employees,
2. Annual net turnover of SEK 350 million,
3. Annual balance sheet of SEK 175 million.

NFRD was transposed as an amendment to Årsredovisningslagen in December 2016, with the mandatory reporting requirement entering into effect for the financial year of 2017 (ÅRL 1995:1554). The study from the Swedish Agency for Growth Policy Analysis on the regulation shows that the companies subjected to the law represent approximately two thirds of Sweden's net turnover and 67% of CO2 emissions within the Swedish business sector (PM 2018:22).

Switzerland is a part of the European Economic Area (EEA), however not a part of the EU and thus not subjected to NFRD. They have not enforced regulation on mandatory reporting of ESG-related issues. Instead, large firms' can voluntarily disclosure their non-financial information. In practice, large companies who choose to report are subjected to the SIX Exchange Regulation and the Directive on Information Relating to Corporate Governance (SIX 2022). It requires the sustainability report to be in accordance with a standard such as the Global Reporting Initiative (GRI) framework. There are other regulations that deal with voluntary

disclosure, e.g., the Swiss Code of Best Practice for Corporate Governance (however none are mandatory or as extensive as the NFRD.

As a note for further research: The EU Commission has proposed to update the NFRD for a Corporate Sustainability Reporting Directive (CSRD) in the financial year of 2023. The directive amends the existing NFRD by extending the scope of companies obliged, audit requirements, reporting standardization and reporting digitization (EU Commission 2021). The current NFRD lacks a standardization of reporting measures where the CSRD aims to cope with this ambiguity by ensuring comparability and more qualitative information regarding firms' sustainability initiatives.

2.2 ESG-ratings Background

2.2.1 ESG Definition

ESG stands for sustainability within the environmental-, social- and governance dimensions of a firm. A total ESG-score is given as an approximation of a firm's entire undertaking with regards to these three dimensions within sustainability, either when comparing within industry or across the entire investment universe. This approximation is based on indicators of different attributes that measure each underlying dimension. Environmental attributes include, but are not limited to, energy consumption, greenhouse gas emissions and water consumption. Social attributes include, but are not limited to, worker rights, safety, diversity, education, labour relations, supply chain standards, community relations and human rights. Governance attributes consider the decision-making process within a company. This includes, but is not limited to, transparency on board compensation, independence of boards and shareholder rights.

2.2.2 Sustainable Investment Strategies

ESG-ratings are directly or indirectly applied as a benchmark criterion that is incorporated in investment decisions in at least four out of the seven most common sustainable investment practices. See Table 9 for a list of these investment strategies. It stresses the importance of assessing the actual quality of ESG-ratings and raters.

2.2.3 ESG-rating Methodologies

The International Organization of Securities Commission highlights a lack of transparency on the methodologies underpinning firms' ESG-ratings as a major issue (IOSCO 2021). A survey answered by 339 experts and investors on the most important factors that determine ESG-rating quality concluded that respondents are most considerate of the quality of methodology and the disclosure of methodology, rather than common usage by investors and other stakeholders (Wong and Petroy 2020).

ESG-rating methodologies are in general based on a diverse set of sources, including the annual report, sustainability report, news and questionnaires. With this data, raters consider different indicators of the most relevant attributes defined by some criterion. The indicators are then categorized into either environmental-, social- and governance and weighted accordingly, before being aggregated into a score used for comparison. Table 10 shows a more detailed description of the methodology for the ESG-raters included in the sample.

2.3 Data Sources

In the previously mentioned survey by Wong and Petroy (2019, 2020), experts and investors were asked to evaluate specific raters in terms of rating quality and usefulness. The survey found that respondents primarily prefer RobecoSAM Corporate Sustainability Assessment, CDP Climate, Water & Forest Scores, Sustainalytics' ESG Risk Ratings and MSCI ESG Ratings. Other raters mentioned include Bloomberg ESG Disclosure Scores, ISS-Oekom Corporate Rating, FTSE Russell's ESG Ratings, ISS QualityScore, EcoVadis CSR Rating, Thomson Reuters ESG Scores and Vigeo Eiris Sustainability Rating. Together, these raters represent a substantial part of the market for ESG ratings (Eccles and Stroehle 2018). To study the effect of a mandatory reporting requirement on the rating disagreement, one would ideally include all the above-mentioned raters in the dataset. However, these sources are generally restricted by licenses.

Nonetheless, we gained access to three of the above-mentioned raters. Through the Bloomberg terminal, we sourced ESG Disclosure Scores for the period from 2007-2020 and RobecoSAM Corporate Sustainability Assessment scores between 2016-2020. The latter will be referred to as S&P Global ESG Scores since S&P Global acquired RobecoSAM in 2019 (S&P Global Press 2019). We used the Thomson Reuters Eikon database to access Thomson Reuters ESG Scores between 2007-2020, formerly known as Asset4 and today commonly referred to as

Refinitiv ESG Score (Berg, Fabisik and Sautner 2021). For simplicity, we refer to the three raters as Bloomberg, Refinitiv and S&P Global throughout the paper. The three chosen data sources have been applied in previous related literature (Berg et al. 2019; Gibson et al. 2020; Billio, Costola, Hristova, Latino and Pelizzon 2020).

2.4 Sample Selection

We pursue a similar sample selection process as Berg et al. (2019). We retrieve corresponding data from each rater we have access to; Bloomberg, Refinitiv and S&P Global. The data includes the aggregated scores for the total-, environmental-, social- and governance dimensions. The difference in differences approach requires historical ESG-data covering a sufficient period before and after the policy change, while Swedish company data is necessary to determine whether firms are subjected to NFRD or not. The ideal sample would consist of overlapping historical ESG-data for the relevant raters and include the NFRD-obliged Swedish firms and publicly listed Swiss firms. However, raters frequently lack overlapping scores for the same firms and period.

To account for the empirical setting requirements, we extract a common sample consisting of Bloomberg and Refinitiv for 2012-2020 which is used in the regressions. In the summary statistics, we add data from S&P Global from 2016-2020. The sample selection process has the following logic: first, we retrieve the maximum historical ESG-data from all available Swedish- and Swiss firms from each given rater. Then, we filter the data for each score dimension by excluding firms not covered by all the raters and exclude firms which lack a full series of ratings between the baseline year and the final year. Specifically, for the Swedish sample size, we exclude firms which are not obliged to the NFRD amendment, see Section 2.1. Table 11 and 12 shows firm names, industry, and ISIN codes for the common sample of Swiss and Swedish firms respectively. Evidently, we have a representative sample for Swedish firms with a significant proportion of the OMXS30.

2.5 Summary Statistics

In this section we describe three different measurements of disagreement based on the methodology of Berg et al. (2020). We present descriptive statistics, pairwise correlations, mean absolute distances and quintile rankings between raters for all score dimensions.

Table 1: Correlation Statistics

This table reports descriptive statistics of the average correlation included in the pre- and post-treatment period between Refinitiv and Bloomberg for the treatment- and control group. The pre-treatment period includes the years 2012-2016 (5) and the post-treatment period includes the years 2017-2020 (4).

Descriptive Statistics	**Dependent Variables**			
	(1)	(2)	(3)	(4)
Score dimensions:	Total	Environmental	Social	Governance
Treatment, pre				
Mean	0.68	0.51	0.59	0.34
Median	0.67	0.51	0.61	0.34
Min	0.63	0.45	0.44	0.28
Max	0.74	0.56	0.7	0.47
N	5	5	5	5
Treatment, post				
Mean	0.59	0.5	0.28	0.18
Median	0.57	0.5	0.25	0.17
Min	0.54	0.43	0.12	0.07
Max	0.67	0.55	0.49	0.31
N	4	4	4	4
Control, pre				
Mean	0.73	0.41	0.69	0.5
Median	0.73	0.41	0.7	0.53
Min	0.7	0.24	0.63	0.26
Max	0.77	0.58	0.76	0.66
N	5	5	5	5
Control, post				
Mean	0.69	0.42	0.61	0.28
Median	0.69	0.42	0.59	0.29
Min	0.63	0.29	0.56	0.18
Max	0.75	0.54	0.69	0.35
N	4	4	4	4

2.5.1 Correlation

Correlation is a relevant measurement for disagreement as it describes the degree to which two raters move in coordination with one another. If two raters rate in the same direction, both raters will have a positive correlation and vice versa. The correlation coefficient is given by,

$$r_{xy_{it}} = \frac{\Sigma(x_{ijt} - \bar{x}_{it})((y_{ijt} - \bar{y}_{it})}{\sqrt{(\Sigma(x_{ijt} - \bar{x}_{it})^2)(\Sigma(y_{ijt} - \bar{y}_{it})^2)}}$$

where $r_{xy_{it}}$ is the correlation coefficient between raters x and y for the ESG-score dimension i in time t. x_{ijt} and y_{ijt} is the ESG-rating i for firm j in time t from rater x and y. $i = 1,2,3,4$ and corresponds to the total-, environmental-, social- and governance score dimensions. $t = 1$ for 2012.

Table 2 shows the historical correlations between ESG-ratings and is divided into two panels. In Panel A, the average correlation for the entire period (2012-2020) for the total score was 0.64 and 0.71 for Sweden and Switzerland respectively. For both countries, the three pillar scores had a lower average correlation for the entire period compared to the total score. Governance had the lowest average correlation of 0.27 and 0.40 for Sweden and Switzerland. For Sweden, the social- and governance scores had higher variation over the period, ranging from 0.12-0.70 and 0.07-0.47 respectively. Switzerland exhibits the largest variation in the environmental- and governance scores, ranging from 0.24-0.58 and 0.18-0.66 respectively. With the addition of a third rater in Panel B, we observe a decrease in the average correlation for all score dimensions. This is partly explained by a large reduction in sample size. The social- and governance scores have the lowest average correlation, at 0.08 and 0.07 respectively. Furthermore, Bloomberg and S&P Global are relatively more correlated to one another. In contrast, Refinitiv and S&P Global have clearly exhibited the lowest correlations with each other. To conclude, we observe a substantial rating disagreement, consistent with prior findings from Berg et al. (2020).

Table 2: Correlation of ESG-ratings

This table reports historical correlations of aggregate ESG-ratings for the total- (Total), environment- (E), social- (S) and governance (G) dimensions. Panel A shows correlations between Bloomberg and Refinitiv for Swedish and Swiss firms between 2012 to 2020. Panel B shows correlations between Bloomberg, Refinitiv and S&P Global for Swedish firms between 2016 to 2020. BB, RE and SP are abbreviations for Bloomberg, Refinitiv and S&P Global, respectively.

Panel A: Correlation Between Bloomberg and Refinitiv

		2012	2013	2014	2015	2016	2017	2018	2019	2020	Average
		RE BB	RE BB	RE BB	RE BB	RE BB	RE BB	RE BB	RE BB	RE BB	RE BB
Sweden	Total*	0.71	0.63	0.66	0.74	0.67	0.67	0.56	0.54	0.57	0.64
	E**	0.55	0.50	0.51	0.56	0.45	0.51	0.55	0.49	0.43	0.51
	S***	0.59	0.61	0.70	0.61	0.44	0.49	0.29	0.22	0.12	0.45
	G*	0.47	0.34	0.28	0.28	0.35	0.17	0.17	0.07	0.31	0.27
Switzerland	Total†	0.70	0.73	0.77	0.73	0.71	0.63	0.68	0.75	0.69	0.71
	E††	0.24	0.58	0.41	0.45	0.38	0.52	0.54	0.31	0.29	0.41
	S†††	0.76	0.69	0.70	0.70	0.63	0.57	0.56	0.69	0.61	0.66
	G†	0.66	0.55	0.53	0.48	0.26	0.24	0.34	0.35	0.18	0.40

* 41 firms included in the sample size
** 37 firms included in the sample size
*** 39 firms included in the sample size
† 26 firms included in the sample size
†† 17 firms included in the sample size
††† 19 firms included in the sample size

Panel B: Correlation Between Bloomberg, Refinitiv and S&P Global

		2016			2017			2018			2019			2020			Average
		RE BB	RE SP	BB SP	RE BB	RE SP	BB SP	RE BB	RE SP	BB SP	RE BB	RE SP	BB SP	RE BB	RE SP	BB SP	
Sweden	Total*	0.51	0.32	0.48	0.39	0.12	0.61	0.13	0.13	0.60	0.32	-0.01	0.59	0.56	0.02	0.33	0.34
	E*	0.64	0.58	0.31	0.66	0.62	0.40	0.66	0.33	0.42	0.68	0.31	0.44	0.82	0.19	0.23	0.49
	S*	0.06	0.00	0.58	-0.16	-0.15	0.60	-0.12	-0.17	0.41	-0.23	-0.13	0.45	-0.15	0.12	0.08	0.08
	G*	0.31	0.16	0.30	-0.02	0.00	0.30	-0.13	-0.03	0.14	-0.26	-0.15	0.04	0.06	0.06	0.32	0.07

* 22 firms included in the sample size

2.5.2 Mean Absolute Distance

Although correlations are intuitive, they may mask firm-specific differences. Furthermore, correlation could be driven by a small difference for all firms, or by strong outliers for a few firms which heavily affects the correlation for the entire sample. To account for this, we evaluate the heterogeneity of disagreement across firms by measuring the mean absolute distance (MAD) to the average rating for each firm using a normalized sample. This firm-specific measure illustrates which firms' raters tend to agree or disagree upon. We normalize

the common sample between 2016-2020 for Bloomberg, Refinitiv and S&P Global. We normalize the sample with the following equation,

$$x_{norm_{ijt}} = \frac{x_{ijt} - x_{min_{it}}}{x_{max_{it}} - x_{min_{it}}}$$

where $x_{norm_{ijt}}$ is the normalized score for ESG-rating i for firm j in time t. $x_{max_{it}}$ and $x_{min_{it}}$ is the maximum and minimum scores within the dataset for each firm and score dimension. We compute the MAD in accordance with Berg et al. (2020),

$$MAD_{ijt} = \frac{\Sigma \left| x_{norm_{ijt}} - \mu_{ijt} \right|}{N}$$

where MAD_{ijt} is the mean absolute distance of the data points from the mean. $x_{norm_{ijt}}$ is the normalized score from a rater for the ESG-rating i for firm j in time t. μ_{ijt} is the mean of the rater normalized scores for firm j. $\Sigma \left| x_{norm_{ijt}} - \mu_{ijt} \right|$ represents the sum of the absolute deviations of the normalized scores from the mean. The sum of the absolute deviations are divided by N which represents the number of raters which have rated firm j.

Figure 1 (and 4, 5, 6) are graphical illustrations of the firm-specific MAD's and are sorted by their average MAD. The figures illustrate the agreement on firms' ESG-scores. Firms such as Swedish Match AB and Elekta AB consistently have relatively high average MADs across the score dimensions, indicating that the rating disagreement is relatively high.

Table 3: Example of ESG-rating Disagreement

The table reports the social scores for Bloomberg, Refinitiv and S&P Global on Swedbank AB, Svenska Cellulosa AB SCA, Swedish Match AB, and Elekta AB.

Company	Bloomberg	Refinitiv	S&P Global
Swedbank AB	58	67	89
Svenska Cellulosa AB SCA	65	88	33
Swedish Match AB	33	85	38
Elekta AB	22	92	78

Social scores from 2020

Figure 1: Mean Absolute Distance (MAD) – Social Score

This figure reports the MAD to the average rating using the normalized common sample of Swedish firms. Firms are sorted from the lowest to highest mean absolute distance. Bloomberg, Refinitiv and S&P Global are plotted in different colours. Figure 1A and 1B reports the MAD for the Social scores in 2016 and 2020 respectively.

Figure 1A: MAD 2016

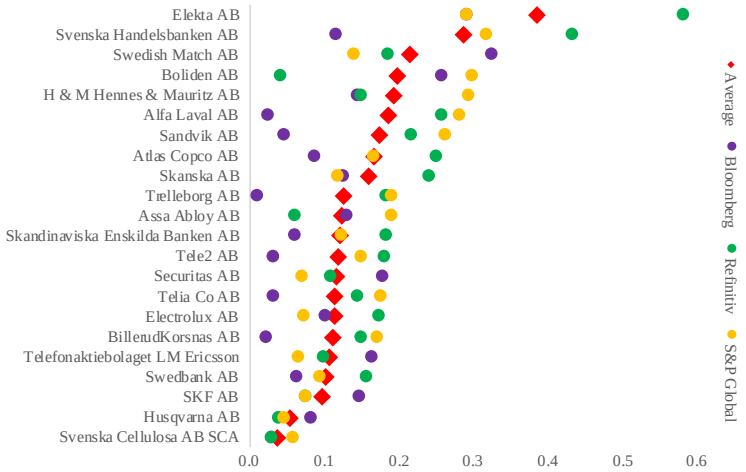


Figure 1B: MAD 2020

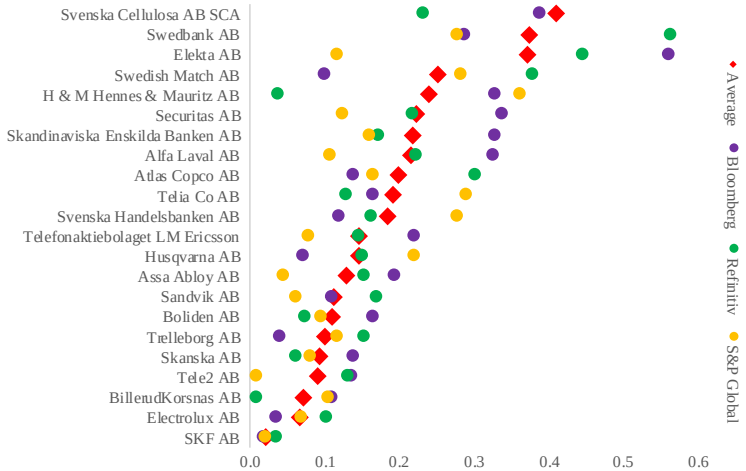


Table 4 shows how the MAD for each score dimension is distributed in 2016 and 2020 between Bloomberg, Refinitiv and S&P Global for Swedish firms. We observe that the maximum MAD has increased across all dimensions, with the largest increase (0.11) for the environmental score, indicating an increased rating disagreement. We also observe that the median, mean and third quartile had increased for the social- and governance scores while decreased for total- and environmental. When comparing the different score dimensions, we observe that the mean and median MADs are similar (ranging from 0.11-0.18) within all dimensions for both years. The mean is slightly greater than the median for both the environmental and social score, indicating a slight positive skewness, while the total- and governance score have no substantial skewness.

Table 4: Mean Absolute Distance (MAD)

These tables report how the MAD to the average rating is distributed in 2016 and 2020. It is based on the normalized common sample of 22 Swedish firms from the three rating agencies, Bloomberg, Refinitiv and S&P Global.

Panel A: Distribution of Total Score Disagreement

		MAD 2016	MAD 2020
Sweden	Minimum	0.01	0.03
	1st Quartile	0.08	0.09
	Median	0.16	0.13
	Mean	0.16	0.14
	3rd Quartile	0.23	0.17
	Maximum	0.31	0.37

Panel B: Distribution of Environment Score Disagreement

		MAD 2016	MAD 2020
Sweden	Minimum	0.04	0.03
	1st Quartile	0.10	0.07
	Median	0.13	0.11
	Mean	0.14	0.13
	3rd Quartile	0.17	0.16
	Maximum	0.28	0.39

Panel C: Distribution of Social Score Disagreement

		MAD 2016	MAD 2020
Sweden	Minimum	0.04	0.02
	1st Quartile	0.11	0.10
	Median	0.12	0.16
	Mean	0.15	0.18
	3rd Quartile	0.18	0.22
	Maximum	0.39	0.41

Panel D: Distribution of Governance Score Disagreement

		MAD 2016	MAD 2020
Sweden	Minimum	0.04	0.06
	1st Quartile	0.12	0.11
	Median	0.16	0.19
	Mean	0.16	0.18
	3rd Quartile	0.21	0.25
	Maximum	0.32	0.34

2.5.3 Ranking Disagreement

Institutional investors frequently screen the worst- or best-in-class companies by their ESG-score as a part of their sustainable investment strategies (Wong and Petroy 2020). For this use case, the level of disagreement measured by correlation or MAD may not be relevant. Instead,

rankings become a more appropriate measure of disagreement. We describe the applied method in text below:

Using the common sample (including and excluding S&P Global) we sort every rater's scores in order from lowest to highest. If any firms overlap amongst the raters in the top- and bottom quintiles (20%) of the sorted datasets, we include those in the table.

The rankings are shown in Table 13 where Panel A includes common quintile rankings between Bloomberg and Refinitiv for Sweden and Switzerland from 2012-2020. Generally, there is a higher consensus regarding the worst-in-class firms within Sweden compared to the best-in-class firms. This suggests which companies one would expect to exclude when screening for sustainable investments. In contrast, Table 14 shows that the consensus is more balanced within Switzerland regarding the best- or worst-in-class firms. Table 5 includes S&P Global yet is limited to Sweden and a shorter period. When comparing Table 13 (two raters) to Table 5 (three raters) for Sweden, there are almost no firms in Table 5 in which all three raters agree to be the best- or worst-in-class. This highlights the prominent issue of ESG-rating disagreement.

Table 5: Quintile Rankings Sweden (BB, RE & SP)

This table reports the top 20% best- or worst-in-class firms for Bloomberg, Refinitiv and S&P Global between 2016 to 2020 in the common sample. See Table 11 for the full firm names.

	2016		2017		2018		2019		2020	
	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile
Total*			BILL				ELUX		ELUX	
E*		TEL2						TEL2		TEL2
S*				TEL2		TEL2				
G*				SKA						

* With 22 firms included in the sample size, 4 firms are included in the top- and bottom quintiles (20%)

3. Research Design

In this section, we present our hypothesis, the ideal empirical setting, and the actual empirical strategy to measure the causal effect of a mandatory sustainability reporting requirement on ESG-rating disagreement.

3.1 Hypothesis

H1: A Mandatory Sustainability Reporting Requirement Decreases ESG-rating Disagreement

The hypothesis is derived from the explicit purpose of NFRD (Directive 2014/95/EU) to increase the transparency, quality, and comparability of information on firms' sustainability performance in the European Union. An important source of ESG-information on which raters base their score is sustainability disclosure in annual- and sustainability reports. For instance, the three ESG-raters in our sample consider disclosed company data in their rating methodologies (S&P Global, Bloomberg, Refinitiv). NFRD should then, logically, decrease the level of measurement error and rating disagreement amongst ESG-raters if it fulfils its desired effects.

3.2 Preliminaries

3.2.1 Comparability Assessment and Limitations in Setting

Within the introduction in the paragraph, Challenge 1: Causal Effect of Mandatory Reporting Requirement, we elaborated on the ideal empirical setting for the hypothesis to be tested. Since this was unattainable, we settled for a somewhat flawed comparability assumption which makes the subsequent test results less reliable and decreases possibility for generalization on the causality. Nevertheless, below we argue that the similarity between Switzerland and Sweden are comparable enough to infer whether NFRD had a significant effect or not.

As illustrated in Table 15, the two countries have similar macroeconomic characteristics. For example, both have a labour force of 5 million, power purchasing parity (Real GDP) of $590 billion and $525 billion respectively and the GDP composition by sector and end use are similar with the most significant deviations relating to government consumption. We find that the average historical rating correlations between Swedish and Swiss firms for Bloomberg and Refinitiv before the implementation of NFRD was on a similar level, see Figure 2. Furthermore, there has been a surge in positive ESG sentiment by Swiss investors in recent years, contrary

to their historical low level of disclosure within the financial sector. This is exemplified by the annual Swiss Sustainable Investment Market Study (Paetzold, Busch, Gerber, Döbeli 2021), which is written by the University of Zurich and Swiss Sustainable Finance, that reports that Switzerland aspires to become a hub for international flows of sustainable finance. Sweden and Switzerland thus share a positive investor sentiment towards ESG.

To continue our analysis, we assume that Sweden and Switzerland are comparable. Yet, limitations of not having the ideal scenario remain. First, the mandatory reporting requirement is only quasi-exogenous since Sweden and Switzerland suffer from different time trends in the development of the ESG-ratings market. This in turn could explain a selection effect, where the fact that Switzerland is not part of the EU may not be random, and thus the mere fact that Switzerland has not been subjected to mandatory reporting requirement decreases the opportunities for causal inference. Second, the selection of sample is restricted to available data on firms. It could be a symptom of omitted variables since firms that are not rated at all, or only by a few raters, might be so due to a non-random characteristic (e.g., a low sustainability performance). Hence, the correlation is naturally biased and skewed towards firms that are rated.

3.2.2 Considerations Regarding Sample Size

As our specific research question has not been explored in prior research to the best of our knowledge, we lack academic reference points to guide us in understanding the appropriate sample size. According to the study made by the Swedish Agency for Growth Policy Analysis (2018), approximately 1,500 Swedish companies were subjected to the NFRD. However, this estimate also includes private firms which are outside the scope of this essay, as their information is frequently unavailable to the public and investors, thus often not covered by ESG-raters. Therefore, we analyse the population size based on all the listed Swedish equities on the Nasdaq Stockholm Stock Exchange, which amounts to 935 shares, of which 406 of the shares are listed on their primary market (Nasdaq OMX Nordic 2022). This includes large-, mid- and small cap companies. Henceforth, the population size for the scope of our paper is Swedish firms obliged to the NFRD amendment with the possibility of being rated. To analyse the second criteria, we retrieve revenue-, balance sheet- and employee data from 2020 to screen according to the NFRD requirements mentioned in Section 2.1. We found 341 listed companies which fulfilled at least two of the three. We apply the same method for Switzerland, which has 250 listed equities on the SIX Swiss Stock Exchange (SIX Swiss Exchange 2022). With the

absence of mandatory reporting requirements, we consider all listed equities as relevant counterfactuals and that there need not be any further delimitation.

Given these population estimates of 341 and 250 companies for Sweden and Switzerland respectively, we simulate an appropriate sample size using the analytical program G*Power. As we are studying the differences in correlation pre- and post a regulatory change between two independent countries, we choose the statistical test; "Means: Difference between two independent means (two groups)". We set a confidence level $(1-\alpha)$ of 95%, a significance level (α) of 5% and a statistical power $(1-\beta)$ of 80%. These measures are set to balance the probability of having a Type I (false positive) and Type II (false negative) error. Given these inputs, the sample size required is 76 and 56 for Sweden and Switzerland respectively. We observe that this might limit the possibility of obtaining significant test results due to our sample size of 41 and 26 for Sweden and Switzerland respectively.

3.3 Empirical Model

Consider an empirical difference in differences model that links the dependent variable $Y_{it} = \text{ESG} - \text{rater Score}_i \text{ Correlation}$ to the independent dummy variable Regulation (Treat x Post) to a treatment- (Sweden) and control group (Switzerland),

$$Y_{it} = \alpha + \beta\,\text{Treatment}_i + \gamma\,\text{Post}_t + \delta\,(\text{Treatment}_i \cdot \text{Post}_t) + \epsilon_{it} \quad (1)$$

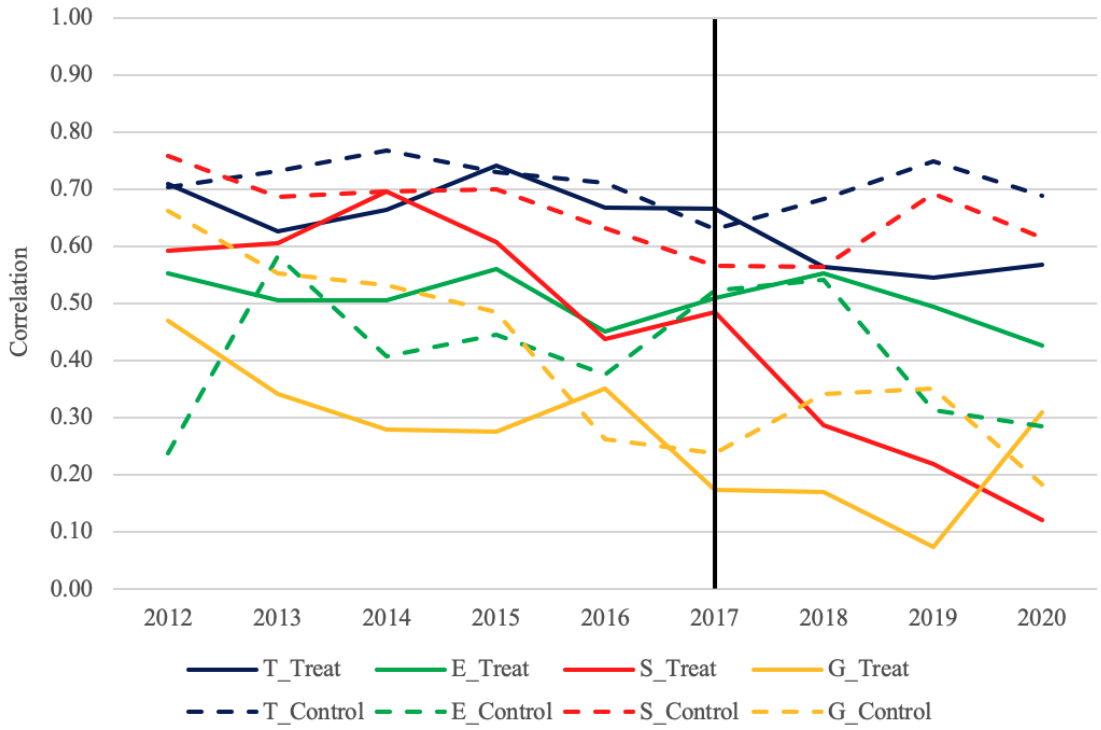
where Treatment_i is a dummy variable for Sweden, Post_t is a dummy variable for whether the NFRD has been implemented or not, where $t = 1$ at and after 2017, and ϵ_{it} is an error term. The coefficient of interest is δ, which is the diff-in-diff estimator and measures the average treatment effect of the regulation. The interpretation of coefficients is listed below:

$\alpha =$ expected value of y (correlation) in period 1 for the control group (Switzerland)
$\alpha + \gamma =$ expected value of y in period 2 for the control group
$\gamma =$ expected change in y from period 1 to period 2 for the control group
$\alpha + \beta =$ expected value of y in period 1 for the treatment group (Sweden)
$\alpha + \beta + \gamma + \delta =$ expected value of y in period 2 for the treatment group
$\beta + \delta =$ expected change in y from period 1 to period 2 for the treatment group
$\delta =$ additional change in y from period 1 to period 2 for the treatment group compared to the control group (referred to as average treatment effect).

The model intends to examine the causal effect of the implementation, by assigning Sweden (who was subjected to NFRD) as treatment group and Switzerland (who was not subjected to NFRD) as the control group. However, the key identifying assumption for this regression model to hold is that the treatment group would have developed with the same trend as the control group if not for the treatment (NFRD regulatory change). The imperfect yet only approximation for this is the historical trend of the two countries, which we have illustrated in Figure 2.

Figure 2: Historical Correlation Trends

This figure reports the historical correlation trends between Refinitiv and Bloomberg for the treatment- and control group. The vertical line in 2017 represents the year NFRD entered into effect.



4. Main Results

In this section, we present the main results of our empirical strategy using the exposure to a mandatory reporting requirement as treatment. Furthermore, this section includes robustness tests.

4.1 The Effect of Sustainability Reporting on ESG Rating Disagreement

We present the estimates from the regression model (1) for all score dimensions. The coefficient δ could ideally be interpreted as the causal effect of a mandatory sustainability reporting requirement on score correlation and ultimately ESG-rating disagreement. The outcomes confirm the general ESG-rating disagreement found in related literature. We contribute with the finding that a mandatory reporting requirement in isolation does not have a substantial causal effect on the ESG-rating disagreement in general. The average treatment effect is visualized in Figure 3.

4.1.1. Outcomes for the Total Score Correlation

Table 6 shows that a mandatory sustainability reporting requirement had a statistically insignificant effect on rating disagreement in general. In column 1, an ordinary least squares (OLS) linear regression was used to test if a mandatory reporting requirement (Treatment x Post) significantly predicted the total score correlation (TSC) for the Refinitiv and Bloomberg ESG-raters. The fitted regression model was equation (1). The overall regression was statistically significant ($\bar{R}^2 = 0.57,\ F(3,14) = 8.491\ , p = 0.002$). Still, we found that the independent variable Treatment x Post did not significantly predict the TSC ($\delta = -0.055,\ p = 0.202$). Column 2 shows the weighted least squares (WLS) linear regression model equivalent, with identical independent- and dependent variables, raters, and equation. The overall regression was statistically significant ($\bar{R}^2 = 0.50,\ F(3,14) = 6.552\ , p = 0.005$). The reporting requirement did not significantly predict TSC ($\delta = -0.055,\ p = 0.225$).

4.1.2 Outcomes for the Environmental Score Correlation

In Table 6 column 3, an OLS linear regression was used to test if the reporting requirement significantly predicted environmental score correlation. The fitted regression model was equation (1). The overall regression was statistically insignificant ($\bar{R}^2 = 0.069,\ F(3,14) = 1.418\ , p = 0.279$). Unsurprisingly, the reporting requirement did not significantly predict ESC ($\delta = -0.025,\ p = 0.793$). Column 4 shows the WLS linear regression equivalent. The

overall regression was statistically insignificant ($\overline{R}^2 = 0.089,\ F(3,14) = 1.552\ , p = 0.245$) and the reporting requirement did not significantly predict ESC ($\delta = -0.025,\ p = 0.796$).

4.1.3 Outcomes for the Social Score Correlation

In Table 6 column 5, an OLS linear regression was used to test if the reporting requirement significantly predicted social score correlation. The fitted regression model was equation (1). The overall regression was statistically significant ($\overline{R}^2 = 0.722,\ F(3,14) = 15.72\ , p = 0.000$). We found that the reporting requirement significantly predicted SSC ($\delta = -0.225,\ p = 0.025$). Column 6 shows the WLS linear regression equivalent. The overall regression was statistically significant ($\overline{R}^2 = 0.6036,\ F(3,14) = 9.727\ , p = 0.001$) and the reporting requirement significantly predicted SSC ($\delta = -0.225,\ p = 0.039$).

4.1.2 Outcomes for the Governance Score Correlation

In Table 6 column 7, an OLS linear regression was used to test if the reporting requirement significantly predicted governance score correlation. The fitted regression model was equation (1). The overall regression was statistically significant ($\overline{R}^2 = 0.5169,\ F(3,14) = 7.064\ , p = 0.004$). However, the reporting requirement did not significantly predict GSC ($\delta = 0.059,\ p = 0.56838$). Column 8 shows the WLS linear regression equivalent. The overall regression was statistically significant ($\overline{R}^2 = 0.517,\ F(3,14) = 7.064\ , p = 0.004$) yet the reporting requirement did not significantly predict GSC ($\delta = 0.059,\ p = 0.554$).

Table 6: Regression Results

This table reports the difference in differences regression results, sorted by the dependent variables total-, environmental-, social- and governance score correlation. Each dependent variable (see Table 16 for definitions) has been regressed with an ordinary- and weighted least squares linear regression.

	Dependent variables							
Pillars:	TSC		ESC		SSC		GSC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Independent variables	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS
Treatment	-0.04715	-0.04715 .	0.104483	0.104483	-0.10657 .	-0.10657*	-0.15549*	-0.15549 .
	(0.02731)	(0.02311)	(0.071493)	(0.062386)	(0.05981)	(0.04086)	(0.06745)	(0.07555)
Post	-0.04159	-0.04159	0.005814	0.005814	-0.08566	-0.08566 .	-0.22048**	-0.22048*
	(0.02897)	(0.02440)	(0.065223)	(0.087699)	(0.06343)	(0.04187)	(0.07154)	(0.07562)
Treatment x Post	-0.05479	-0.05479	-0.024717	-0.024717	-0.22448*	-0.22448*	0.05911	0.05911
	(0.04096)	(0.04292)	(0.092239)	(0.093953)	(0.08971)	(0.09853	(0.10117)	(0.09742)
Intercept	0.7292***	0.7292***	0.4103***	0.4103***	0.6947***	0.6947*	0.4988***	0.4988***
	(0.01931)	(0.01408)	(0.043482)	(0.059747)	(0.04229)	(0.02147)	(0.04769)	(0.05896)
F-statistic	8.491	6.552	1.418	1.552	15.72	9.727	7.064	7.14
Adj. R-squared	0.5693	0.4949	0.06865	0.08875	0.722	0.6036	0.5169	0.52
p-value	0.002	0.005	0.279	0.245	0.000	0.001	0.004	0.004
n	18	18	18	18	18	18	18	18

. Significant at 10 %, *Significant at 5 %, **Significant at 1%, ***Significant at 0.1%. Standard errors reported in parenthesis.

4.2 Robustness tests

We conduct two different robustness tests to investigate whether our results in the regression was due to noise in the data or a subject of omitted variables.

4.2.1 Placebo Regulation Regression

One way to test the robustness of our model is to estimate a regression that examines the treatment impact before the treatment has occurred, which is supposed to be zero. Consider the same fitted regression model, equation (1), with an adjustment of the implementation year of NFRD from 2017 to 2014. The first year remains 2012 while ending 2016 to avoid including effects from the actual NFRD. We keep the independent variables and test them on TSC and SSC, since they exhibit the highest $\overline{R}^2$. Hence, we have identical treatment and control groups. Table 7 illustrates the insignificance of all placebo regulation regressions for both variables, which was expected.

Table 7: Placebo Regression Results

This table reports the difference in differences regression results that examines the effect of the treatment before it occurred, also called the Placebo regression. It is sorted by the dependent variables total- and social score correlation. The Post dummy variable = 1 for 2014-2016 (see Table 16 for definitions). Each dependent variable has been regressed with an ordinary- and weighted least squares linear regression.

	Dependent variables			
Pillars:	TSC		SSC	
	(1)	(2)	(3)	(4)
Independent variables	OLS	WLS	OLS	WLS
Treatment	-0.049328	-0.049328	-0.12272	-0.12272
	(0.039278)	(0.046841)	(0.08172)	(0.07174)
Post	0.018992	0.018992	-0.04535	-0.04535
	(0.035856)	(0.027417)	(0.07460)	(0.04250)
Treatment x Post	0.003626	0.003626	0.02691	0.02691
	(0.050708)	(0.055706)	(0.10550)	(0.09912)
Intercept	0.717817***	0.717817***	0.72189***	0.72189***
	(0.027774)	(0.023595)	(0.05779)	(0.02605)
F	1.427	1.412	1.561	2.26
Multiple R-squared	0.4164	0.4138	0.4383	0.5305
Adjusted R-squared	0.1246	0.1207	0.1575	0.2957
p-value	0.3245	0.3283	0.2937	0.1818
n	10	10	10	10

. Significant at 10 %, *Significant at 5 %, **Significant at 1%, ***Significant at 0.1%. Standard errors reported in parenthesis.

In Table 7 column 1, an OLS linear regression was used to test if a placebo reporting requirement significantly predicted TSC. The fitted regression model was equation (1). The overall regression was not statistically significant ($\overline{R}^2 = 0.121,\ F(3,14) = 1.427\ , p = 0.325$). As predicted, the placebo reporting requirement did not significantly predict TSC ($\delta = 0.003,\ p = 0.945$). Column 2 shows the WLS linear regression equivalent. The overall regression was not statistically significant ($\overline{R}^2 = 0.121,\ F(3,14) = 1.412\ , p = 0.3283$). Continuing, the reporting requirement did not significantly predict TSC ($\delta = 0.004,\ p = 0.950$).

In Table 7 column 3, an OLS linear regression was used to test if a placebo reporting requirement significantly predicted SSC. The fitted regression model was equation (1). The

overall regression was not statistically significant ($\overline{R}^2 = 0.1575,\ F(3,14) = 1.561\ , p = 0.294$). As predicted, the placebo reporting requirement did not significantly predict SSC ($\delta = 0.0269,\ p = 0.807$). Column 4 shows the WLS linear regression equivalent. The overall regression was not statistically significant ($\overline{R}^2 = 0.2957\ F(3,14) = 2.26\ , p = 0.1818$). Continuing, the reporting requirement did not significantly predict SSC ($\delta = 0.0269,\ p = 0.795$). Note, while reducing the period from 9 to 5 years, the corresponding sample size is reduced proportionally. It could be a contributing factor to the insignificant results.

4.2.2 Randomized Sample Regression

To examine whether the effects are due to a skewed sample with strong outliers in either treatment or control, we do a second robustness test. Consider an identical regression model, equation (1). We use the common sample for the TSC, and SSC respectively, then randomly assigns a country to each firm and divide the firms into the same proportions as in the original sample. We denote the larger sample as "Sweden" and the smaller sample as "Switzerland". We then rerun the same regressions on the new data. In this robustness test we get significant coefficients for the TSC at the 10 % level, which was unexpected, yet does not have a substantial impact on our analysis since TSC original regression results were insignificant for the coefficient of interest. We receive insignificant results for the SSC variable.

In Table 8 column 1, an OLS linear regression was used to test if a reporting requirement significantly predicted TSC with a randomized sample. The fitted regression model was equation (1). The overall regression was statistically significant ($\overline{R}^2 = 0.830,\ F(3,14) = 28.56\ , p = 0.000$). The reporting requirement with the randomized sample significantly predicted TSC ($\delta = 0.078,\ p = 0.052$). Column 2 shows the WLS linear regression equivalent. The overall regression was statistically significant ($\overline{R}^2 = 0.831,\ F(3,14) = 28.89\ , p = 0.000$). Continuing, the reporting requirement with randomized sample significantly predicted TSC ($\delta = 0.078,\ p = 0.052$).

In Table 8 column 3, an OLS linear regression was used to test if a reporting requirement with the randomized sample significantly predicted SSC. The fitted regression model was equation (1). The overall regression was statistically significant ($\overline{R}^2 = 0.694,\ F(3,14) = 13.83\ , p = 0.000$). On the contrary, the randomized sample regression did not significantly predict SSC ($\delta = -0.003,\ p = 0.952$). Column 4 shows the WLS linear regression equivalent. The

overall regression was statistically significant ($\overline{R}^2 = 0.681$ $F(3,14) = 13.11$, $p = 0.000$). Still, the reporting requirement did not significantly predict SSC ($\delta = -0.003$, $p = 0.951$).

Table 8: Randomized Sample Regression Results

This table reports the difference in differences regression results that examines the effect of the treatment with a randomized sample (see Table 16 for definitions). It is sorted by the dependent variables total- and social score correlation. Each dependent variable has been regressed with an ordinary- and weighted least squares linear regression.

	Dependent variables			
Pillars:	TSC		SSC	
	(1)	(2)	(3)	(4)
Independent variables	OLS	WLS	OLS	WLS
Treatment	-0.17927***	-0.17927***	-0.110795**	-0.110795**
	(0.02430)	(0.02419)	(0.036505)	(0.032047)
Post	-0.11441***	-0.11441***	-0.127501**	-0.127501**
	(0.02577)	(0.02577)	(0.038720)	(0.035149)
Treatment x Post	0.07752 .	0.07752 .	-0.003419	-0.003419
	(0.03645)	(0.03645)	(0.054758)	(0.055627)
Intercept	0.89945***	0.89945***	0.792569***	0.792569***
	(0.01718)	(0.01690)	(0.025813)	(0.019884)
F	28.56	28.89	13.11	13.83
Multiple R-squared	0.8596	0.8609	0.7374	0.7477
Adjusted R-squared	0.8295	0.8311	0.6811	0.6936
p-value	0.000	0.000	0.000	0.000
n	18	18	18	18

. Significant at 10 %, *Significant at 5 %, **Significant at 1%, ***Significant at 0.1%. Standard errors reported in parenthesis.

In Table 8 column 3, an OLS linear regression was used to test if the randomized sample significantly predicted SSC. The fitted regression model was equation (1). The overall regression was not statistically significant ($\overline{R}^2 = 0.1575$, $F(3,14) = 1.561$, $p = 0.294$). The randomized sample did not significantly predict SSC ($\delta = 0.0269$, $p = 0.807$). Column 4 shows the WLS linear regression equivalent. The overall regression was not statistically significant ($\overline{R}^2 = 0.2957$ $F(3,14) = 2.26$, $p = 0.1818$). Continuing, the reporting requirement with the randomized sample did not significantly predict SSC ($\delta = 0.0269$, $p = 0.795$).

5. Mechanisms

In this section, we summarize the empirical and sample-related limitations and provide mechanisms which could explain our findings.

5.1 Empirical and Statistical Limitations

The short available period and the annual intervals of data, in combination with a lack of access to all relevant ESG-raters' historical data, pose a considerable limitation to the findings. Furthermore, the actual empirical setting deviates substantially from the ideal. Limitations on the assumption of true comparability between Sweden and Switzerland decreases the possibility for a robust application of the difference in differences model. Nevertheless, we consider our findings a starting point for further research on how legal policies affect the measurement accuracy of ESG-raters.

5.2 Reported Indicators Decrease

The study by the Swedish Agency for Growth Policy Analysis (2018) could be of relevance in explaining our findings. Using data from the Nordic Compass database between 2014-2017, they investigate whether Swedish companies' sustainability reporting has become more transparent and comparable over time. Their findings do not show any increase in reported information over time, instead they observe a slight decrease in the average amount of reported indicators. Despite their study only including one year (2017) of data under the statutory requirements of the NFRD, a replication of their study using data of reported indicators prolonged to 2020 could potentially support our findings for the level of rating disagreement. An improvement to our study would be to integrate the changes in firms' reported indicators over time as controlled variables in the regression model. This addition could potentially explain the incremental impact of specific indicators on correlation, which relates to the concept of materiality. Unfortunately, this is impossible with a difference in differences approach since Switzerland does not have a similar database.

5.3 Silence is Safest

The paper of Bond and Zeng (2021) provides another possible explanation for why the results are insignificant in general. Their study captures the notion that a firm may prefer to stay silent i.e., not disclose ESG-information, since any disclosure will be perceived as negative for some audiences, while staying silent avoids this reputational risk. As the NFRD requires firms to

publish information related to ESG-issues and lack standardized indicators and audit processes (see Section 2.1), it allows firms to circumvent reporting certain unfavourable yet material ESG-related indicators. Thus, the opportunity to stay silent while being subjected to a broadly defined reporting requirement does not lead to a higher degree of informational quality and transparency for ESG-raters. Therefore, such a policy would have a marginal effect on rating disagreement. Instead, a more specific policy design including standardized and audited indicators (e.g., the Corporate Sustainability Reporting Directive proposal) in the non-financial reporting requirement may lead to different results.

5.4 Revised Rating Methodologies

Methodologies are subjected to change over time as raters integrate new criterions into their assessments in response to stakeholders' shifting preferences. The rapid development of new global challenges creates an even more complex rating environment. Putting this phenomenon into the context of this thesis, a methodology change is difficult to anticipate due to the general lack of transparency of when these occur. However, a study by Berg, Fabisik and Sautner (2021) revealed widespread changes when comparing a sample of firms historical ESG-scores based on two data versions: before- and after a change in methodology by Refinitiv. While Refinitivs' old methodology cannot be accessed, Berg et al. (2021) observed that 13% of the scores received rating upgrades while 87% received downgrades. Still, it is not apparent why the ratings were changed and according to which criteria. These sudden and potentially more frequent changes in ESG-scores could give one explanation for why the ESG-rating disagreement has increased in recent years. Furthermore, while assuming that the rating changes are a competitional aspect of differentiating their services and improving their measurement, these score changes could in practice create scepticism for investors regarding the credibility of current methodologies applied by raters.

6. Conclusion

This thesis investigates whether a mandatory sustainability reporting requirement has a causal effect on ESG-rating disagreement, as measured by correlation across ESG-raters. We exploit the Non-Financial Reporting Directive (2014/95/EU) as a quasi-natural experiment on the assignment of a reporting requirement to countries. We apply a difference in differences approach on a linear regression model where we assume comparability of Swedish and Swiss firms. We find that the reporting requirement did not significantly predict a change in rating disagreement when measured by the total-, environmental- and governance scores. Contrary to our hypothesis, point estimates were negative for all but governance. We find that the reporting requirement significantly predicted a change in the social score correlation with a negative coefficient of -0.23, indicating an increased rating disagreement after the introduction of the directive. While the robustness tests applied to the social score correlation were insignificant, we interpret this effect conservatively with limited ability to draw generalizable conclusions. The empirical model leaves room for omitted variables while simultaneously covering an event time with substantial changes in the external environment (e.g., the surge of financial interest in ESG-issues) of the studied phenomenon, increasing the uncertainty of whether NFRD caused the observed variation. To conclude, a broadly defined sustainability reporting requirement appears to have a negligible effect on rating disagreement.

References

Avramov, Doron, and Si Cheng, Abraham Lioui, Andrea Tarelly, 2021, Sustainable investing with ESG rating uncertainty, Journal of Financial Economics, corrected proof.

Berg, Florian, Julian Koelbel, and Roberto Rigobon, 2019, Aggregate Confusion: The Divergence of ESG Ratings, Review of Finance, forthcoming.

Berg, Florian, Julian Koelbel, Anna Pavlova, and Roberto Rigobon, 2021, ESG Confusion and Stock Returns: Tackling the Problem of Noise

Berg, Florian, Kornelia Fabisik, Zacharias Sautner, 2021, Is History Repeating Itself? The (Un)predictable Past of ESG Ratings. Finance Working Paper N 708/2020, European Corporate Governance Institute.

Bénabou, Roland, and Jean Tirole, 2010, Individual and Corporate Social Responsibility, . Economica 77, 1-19.

Billio, Monica, Michele Costola, Iva Hristova, Camelo Latino, and Loriana Pelizzon, 2020, Inside the ESG Ratings: (Dis)Agreement and Performance, SAFE Working Paper No. 284.

Bloomberg, 2022, Global Environmental, Social and Governance – ESG data. Retrieved 2022, April 5 from Bloomberg Web Site: https://www.bloomberg.com/professional/dataset/global-environmental-social-governance-data/

Bond, Philip, Yao Zeng, 2021, Silence is safest: Information disclosure when the audience's preferences are uncertain, Journal of Financial Economics, corrected proof.

Central Intelligence Agency (CIA), 2021, The World Factbook: Sweden, October.

Central Intelligence Agency (CIA), 2021, The World Factbook: Switzerland, October.

Eccles, Robert G., Judith Stroehle, 2018, Exploring Social Origins in the Construction of ESG Measures, SSRN, University of Oxford – Said Business School.

European Commission, COM/2021/189, Proposal for a Corporate Sustainability Reporting Directive (CSRD).

European Parliament Directive 2014/95/EU (NFRD).

Freeman, Edward, 2010, Strategic Management: A Stakeholder Approach (Cambridge University Press).

Friedman, Milton, 1970, A Friedman doctrine - the Social Responsibility of Business Is to Increase Its Profits, The New York Times, September 13.

Gensler, Gary, 2021, Securities and Exchange Commission Speech: Prepared Remarks Before the Principles for Responsible Investment "Climate and Global Financial Markets", July 17.

Gibson, Rajna, Philipp Krueger, and Peter Schmidt, 2021, ESG Rating Disagreement and Stock Returns, Financial Analyst's Journal 74, 104-127.

Hand, Dean, Sophia Sunderji, Noshin Nova, Indrani De, 2021, Impact Investing Decision Making: Insights on Financial Performance, Global Impact Investing Network (GIIN), p.2.

Hartzmark, Samuel, and Abigail Sussman, 2019, Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows. Journal of Finance 74, 2789-2837.

International Organization of Securities Commission (IOSCO), 2021, Environmental, Social and Governance Ratings and Data Products Providers Final Report.

Liang, Hao, and Luc Renneboog, 2017, On the Foundations of Corporate Social Responsibility, Journal of Finance 72, 853-910.

Lins, Karl, Henri Servaes, and Ane Tamayo, 2017, Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis. Journal of Finance 72, 1785 - 1823.

Paetzold, Falko, Timo Busch, Jean-Daniel Gerber, and Sabine Döbeli, 2021, Swiss Sustainable Investment Market Study 2021, University of Zurich - Center for Sustainable Finance and Private Wealth in collaboration with Swiss Sustainable Finance.

Pedersen, Lasse H., Shaun Fitzgibbons and Lukasz Pomorski, 2021, Responsible investing: The ESG-efficient frontier, Journal of Financial Economics 142, 572-597.

Principles for Responsible Investment (PRI), 2021, Annual Report.

Refinitiv, 2022, Environmental, Social and Governance Scores from Refinitiv, retrieved April 5 from Refinitiv Web Site:
https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf

Standard and Poor Global, 2019, Guide to Credit Rating Essentials: What are Credit Ratings and how do they work?

Swedish Government, Department of Justice, "Annual Reporting Law" - Årsredovisningslag (ÅRL), 1995:1554

Swedish Ministry of Finance, 2021, Memorandum: A new objective for the sixth AP fund investment practice, Fi2021/03315

S&P Global, 2019, S&P Global to Acquire the ESG-ratings Business from RobecoSam, New York and Zurich press release Nov 21.

S&P Global, 2022, S&P Global ESG Scores methodology, retrieved April 5 from S&P Global Web Site:
https://www.spglobal.com/esg/documents/sp-global-esg-scores-methodology-2022.pdf

SIX Swiss Exchange on Sustainability Reporting. Retrieved 2022 May 2 from SIX Exchange Web Site:
https://www.six-group.com/en/products-services/the-swiss-stock-exchange/market-data/shares/sustainability-reporting.html

The Swedish Agency for Growth Policy Analysis, Promemoria 2018:022, From Voluntary to Mandatory Sustainability Reporting, 2018/070.

Wong, Christina, and Erika Petroy, 2020, Rate the Raters 2020: Investor Survey and Interview Results.

Wong, Christina, Erika Petroy, and Aiste Brackley, 2019, Rate the Raters 2019: Expert Views on ESG-raters.

7. Figures and Tables

Figure 3: Average Treatment Effect

These figures report the average treatment effect, δ, for the total-, environmental-, social- and governance score dimensions. The pre-treatment period represents the average correlation for the treatment- and control group between 2012-2016 and the post-treatment period represents the average correlation for the treatment- and control group between 2017-2020. The treatment represents the NFRD implementation. The counterfactual illustrates the approximated correlation post-treatment if the treatment group had not been subjected to the treatment.

Figure 3A: Total Score Correlation

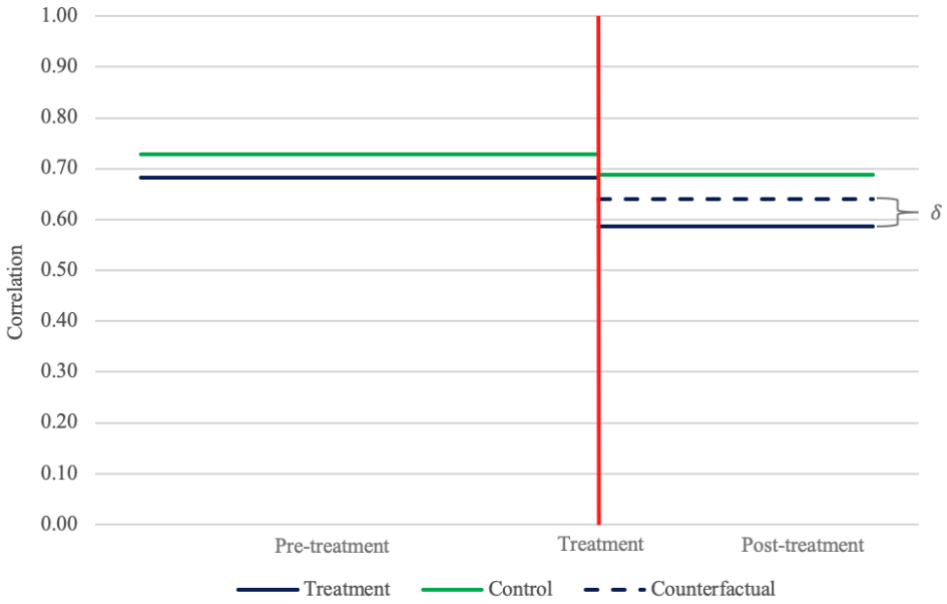


Figure 3B: Environmental Score Correlation

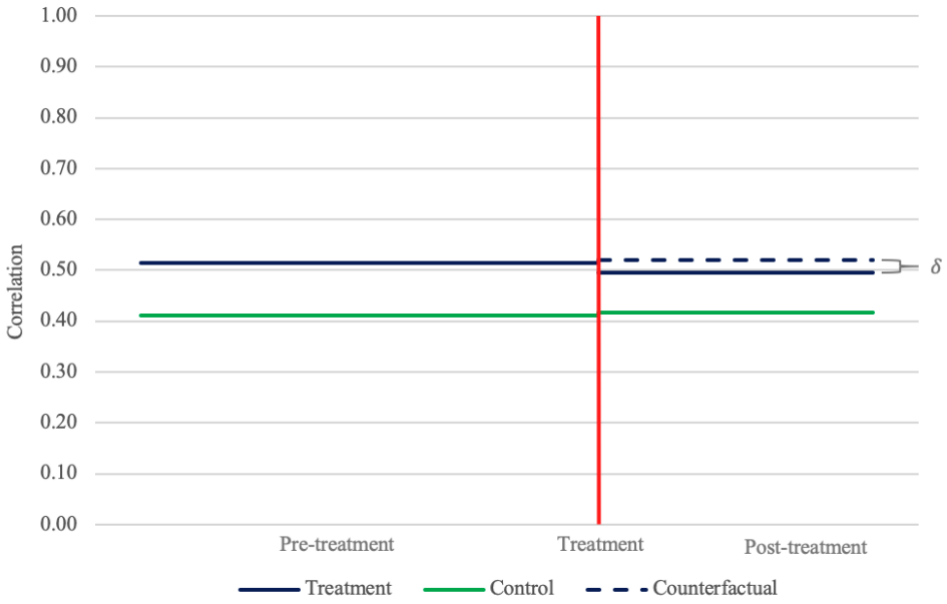


Figure 3C: Social Score Correlation

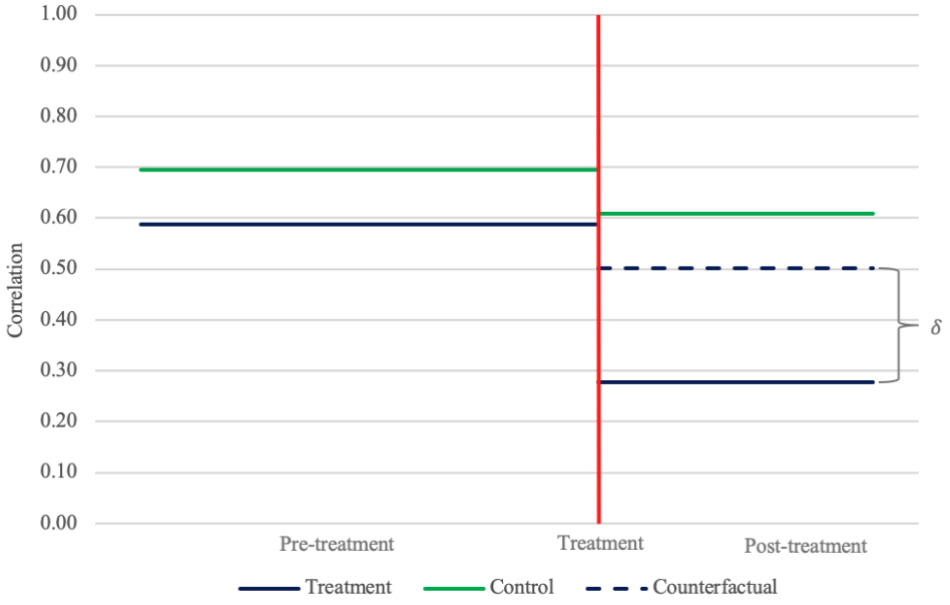


Figure 3D: Governance Score Correlation

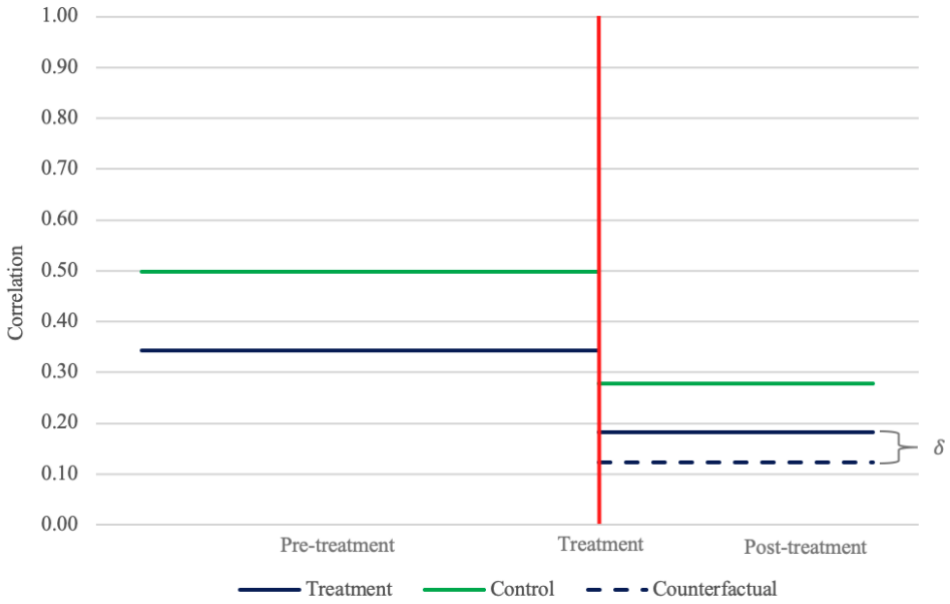


Figure 4: Mean Absolute Distance (MAD) – Total Score

These figures report the MAD to the average rating using the normalized common sample of Swedish firms. Firms are sorted from the lowest to highest mean absolute distance. Bloomberg, Refinitiv and S&P Global are plotted in different colours. Figure 4A and 4B reports the MAD for the Total scores in 2016 and 2020 respectively.

Figure 4A: MAD 2016

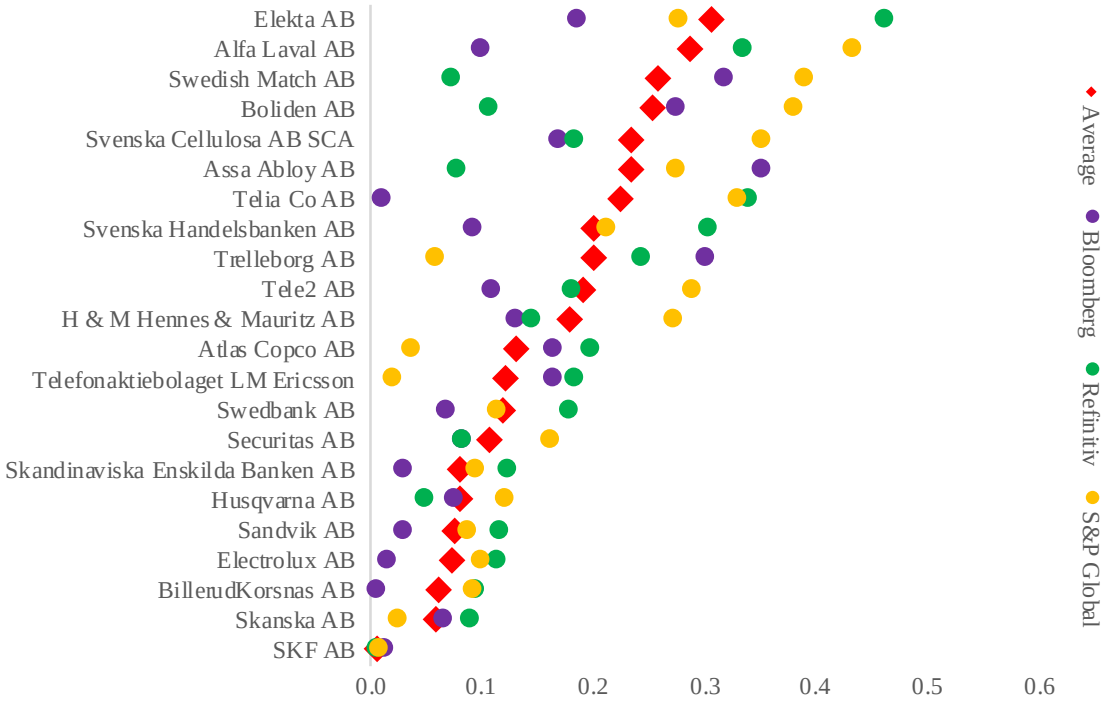


Figure 4B: MAD 2020

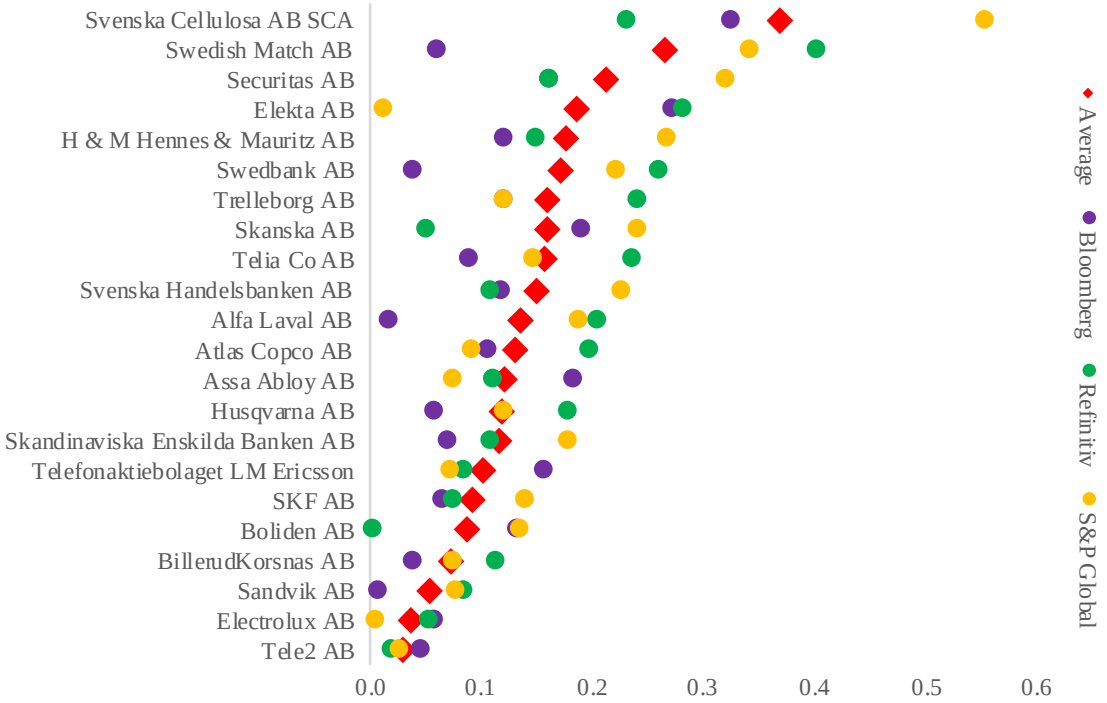


Figure 5: Mean Absolute Distance (MAD) – Environment Score

This figure reports the MAD to the average rating using the normalized common sample of Swedish firms. Firms are sorted from the lowest to highest mean absolute distance. Bloomberg, Refinitiv and S&P Global are plotted in different colours. Figure 5A and 5B reports the MAD for the Environment scores in 2016 and 2020 respectively.

Figure 5A: MAD 2016

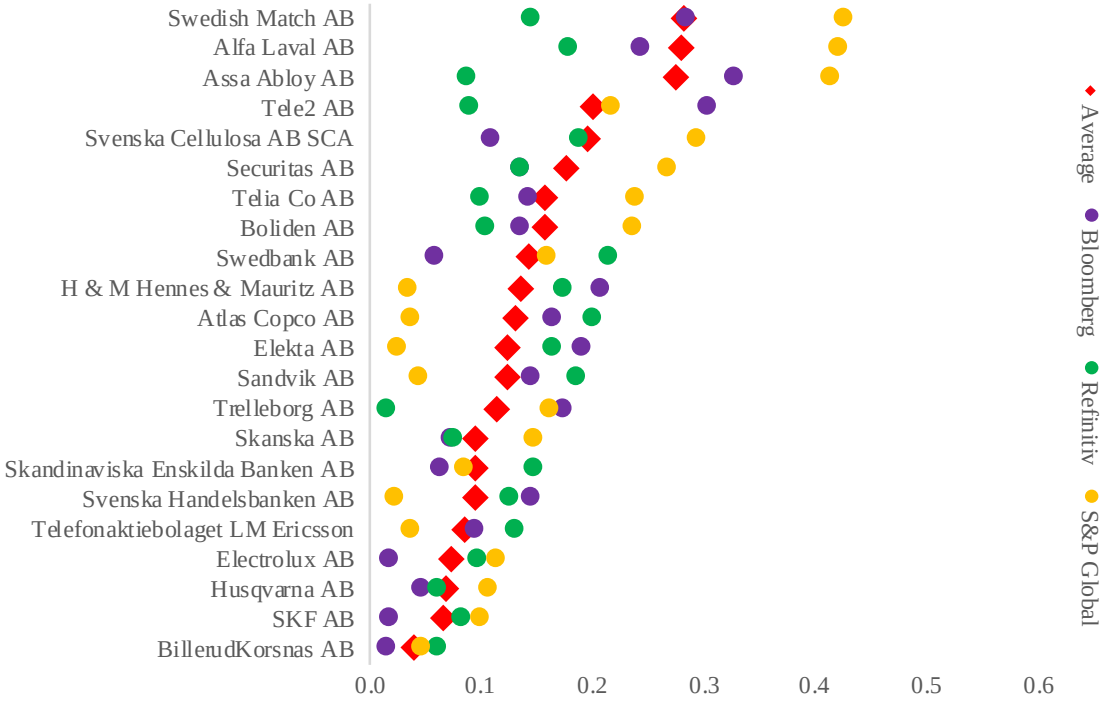


Figure 5B: MAD 2020

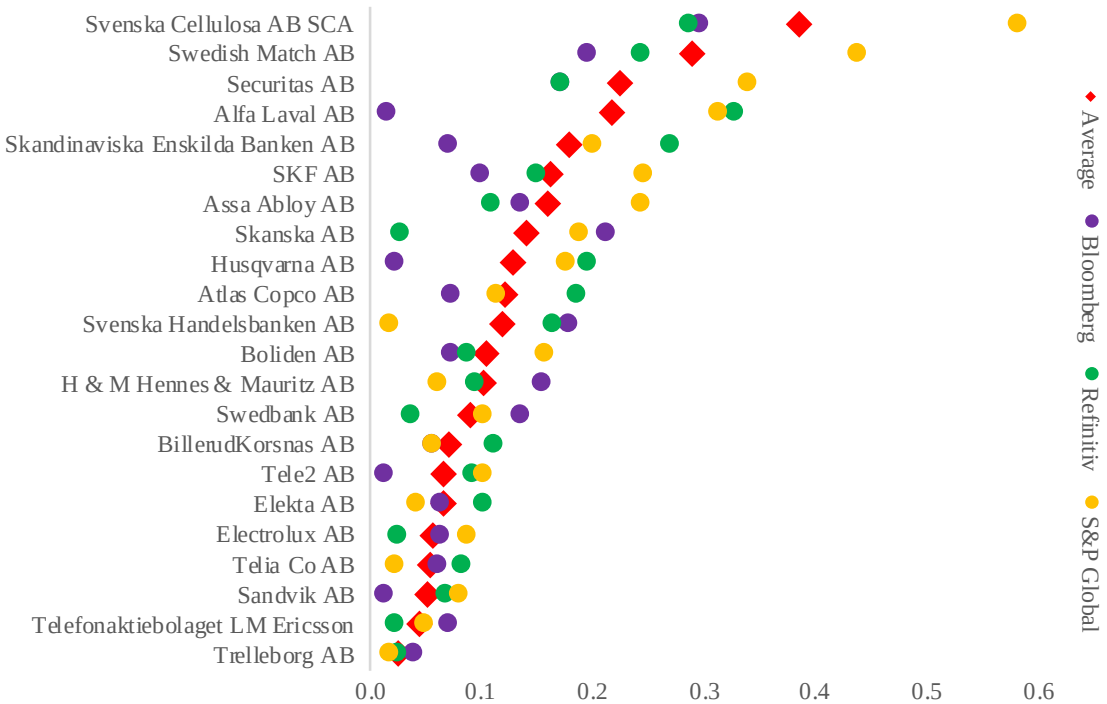


Figure 6: Mean Absolute Distance (MAD) – Governance Score

This figure reports the MAD to the average rating using the normalized common sample of Swedish firms. Firms are sorted from the lowest to highest mean absolute distance. Bloomberg, Refinitiv and S&P Global are plotted in different colours. Figure 6A and 6B reports the MAD for the Governance scores in 2016 and 2020 respectively.

Figure 6A: MAD 2016

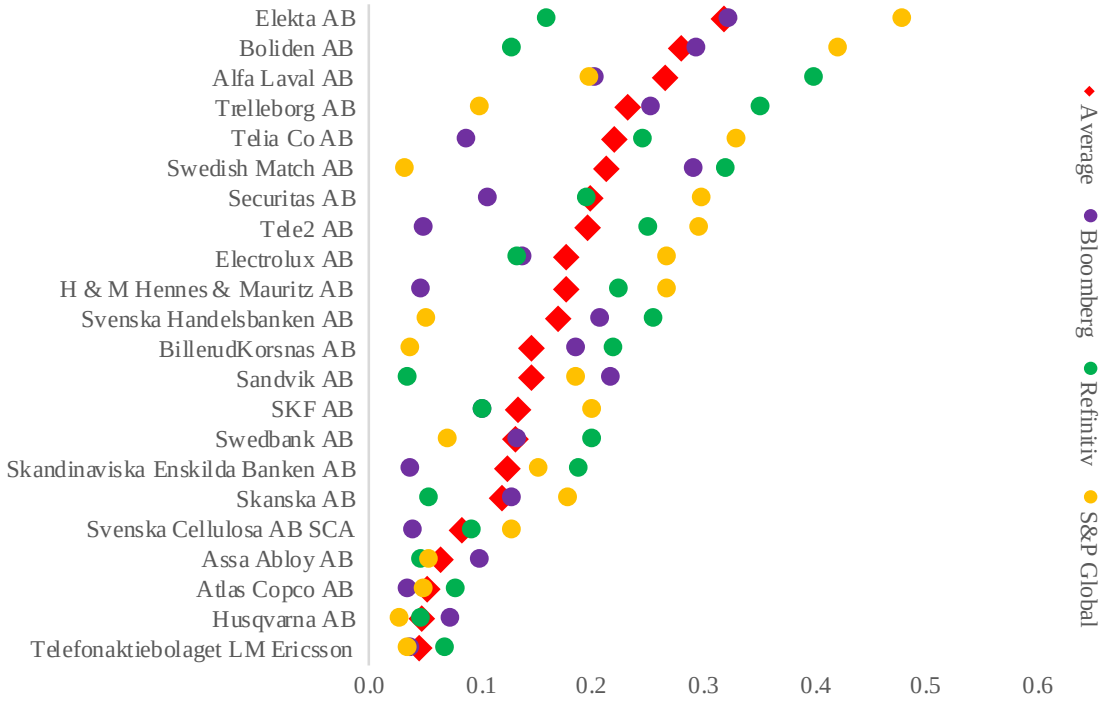


Figure 6B: MAD 2020

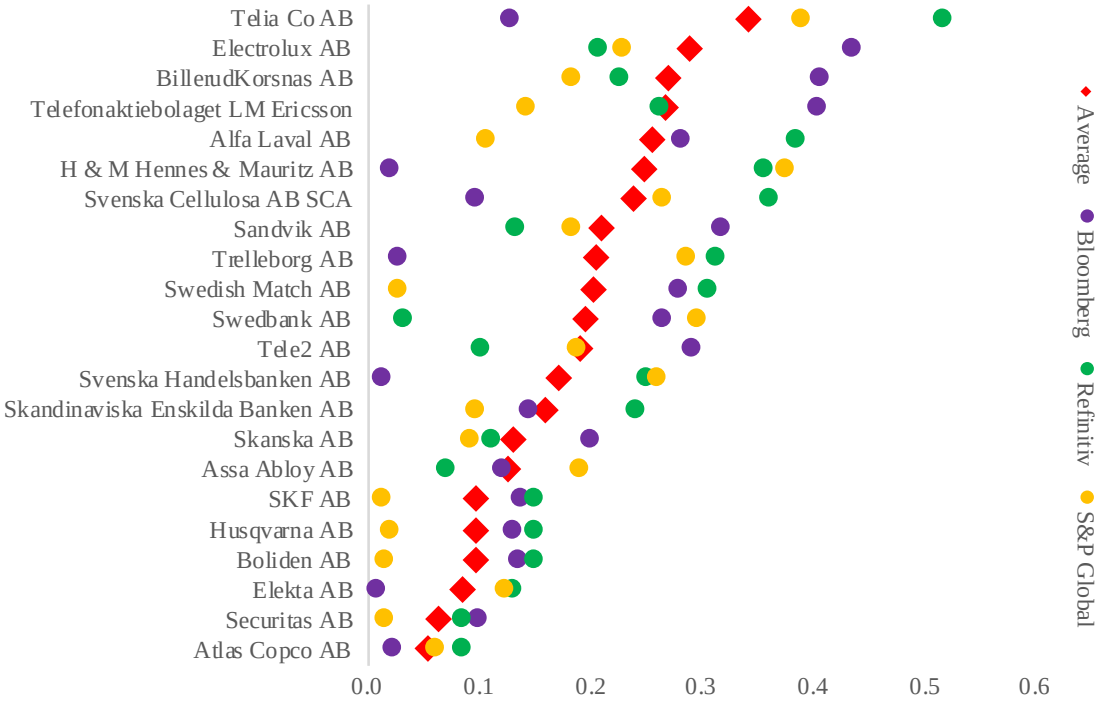


Table 9: Sustainable Investment Strategies

This table reports commonly applied sustainable investment strategies.

Sustainable Investment Strategies	Description
ESG Integration	A systematic inclusion of ESG risk factors in conventional investing activities to maximize financial return, where ESG creates risks and opportunities.
ESG Engagement	Shareholder attempts to influence management to incorporate ESG criteria's within the management of the firm, including but not limited to, filing shareholder proposals and dialogues with goal of changing the company's strategy.
Exclusions (Negative Screening)	Excluding companies, countries or issuers based on a specific exclusion criterion, which can be based on norms and value, product categories (e.g. weapons, adult films), activities (e.g. animal testing) or business practices (e.g. violation of human rights, corruption).
Norms-based Screening	Subcategory of exclusions where the negative screening criteria is commonly based on adhering to certain international norms, such as the ones presented by the International Labour Organization or similar.
Best-in-Class Screening (Positive Screening)	Comparison of ESG performance against its peers (i.e. the same industry or similar categorization) based on ESG rating, often on a threshold basis.
Impact Investing	The Global Impact Investing Network (GIIN) defines Impact investing as: "Investments made with the intention to generate positive, measurable social and environmental impact alongside a financial return". These investments can have expected returns above, below or in line with conventional benchmarks.
Sustainable Thematic Investment	Investment in firms that contributes to solving sustainability-linked issues, both in the social and environmental aspects of sustainability. E.g. renewable energy, energy efficiency, clean technology, low-carbon transportation infrastructure, water treatment and resource efficiency within environmental. Within social sustainability, it regards issues such as education, health systems, poverty reductions and solutions for an aging society.

Table 10: ESG-rating Methodology

This table reports a summary of the methodology used by a few selected ESG-raters.

ESG-rater	Summarised Methodology
Thomson Reuters ESG Scores	Thomson Reuters ESG Scores provides ESG-data for over 12,000 companies and covers over 80% of global market capitalization. Thomson Reuters uses a 0-100 rating scale and sources firm data from annual reports, company websites, NGO websites, stock exchange filings, CSR reports and news sources. Refinitiv calculates over 630 company-level ESG-measures, of which 186 of the most comparable and material per industry are chosen to power the overall company assessment and scoring process. The subset of measures are grouped into categories which reformulate into a environmental-, social-, governance- and total score.
RobecoSAM Corporate Sustainability Assessment	RobecoSAM Corporate Sustainability Assessment provides ESG-data for over 11,500 companies and covers approximately 99% of global market capitalization. RobecoSAM uses a 0-100 rating scale and sources firm data from web-based questionnaires and company documents. RobecoSAM considers approximately 1,000 data points to answer 130+ questions. These questions are weighted into 30+ criteria scores, of which are weighted into environmental-, social- and governance dimension scores. The sum of the weighted dimension scores give the overall ESG-score.
Bloomberg ESG Disclosure Scores	Bloomberg ESG Disclosure Scores provides ESG-data for over 11,500 companies and covers approximately 88% of global market capitalization. Bloomberg uses a 0-100 rating scale. Bloomberg does not publicly disclose their methodology in detail, thus we can not access their aggregation method.
MSCI ESG Ratings	MSCI ESG Ratings is another major rater which provides ESG-data for over 9,000 companies worldwide. MSCI uses a CCC-AAA rating scale and sources firm data from specialized datasets, company disclosure and monitored media sources. The model evaluates data points across 35 ESG key issues that are material for the industry. Then, weights are determined for each key issue based on the firms' risk exposure and risk management, to then derive an environment-, social-, governance dimension score.

Table 11: Sweden Sample

This table reports industry names, SIC codes, tickers and ISIN codes for all Swedish firms included in common sample size 1 with corresponding ratings across Bloomberg and Refinitiv between 2012-2020.

Sample Statistics	(1)	(2)	(3)	(4)
Firm Name	Industry Name	SIC	Ticker	ISIN Code
Elekta AB (publ)	Engineering and Management Services	87	EKTA-B.ST	SE0000163628
SAS AB	Transportation by Air	45	SAS.ST	SE0003366871
Atlas Copco AB	Wholesale Trade Durable Goods	50	ATCO-A.ST	SE0011166610
Castellum AB	Real Estate	65	CAST.ST	SE0000379190
Axfood AB	Food Stores	54	AXFO.ST	SE0006993770
Assa Abloy AB	Business Services	73	ASSA-B.ST	SE0007100581
Nobia AB	Electronic and Other Electric Equipment	36	NOBI.ST	SE0000949331
BillerudKorsnäs AB (publ)	Paper and Allied Products	26	BILL.ST	SE0000862997
Modern Times Group MTG AB	Communication	48	MTGB.ST	SE0000412371
Fabege AB	Real Estate	65	FABG.ST	SE0011166974
Hexagon AB	Business Services	73	HEXA-B.ST	SE0015961909
Investor AB	Security and Commodity Brokers	62	INVEB.ST	SE0015811963
Getinge AB	Instruments and Related Products	38	GETI-B.ST	SE0000202624
Husqvarna AB	Industrial Machinery and Equipment	35	HUSQB.ST	SE0001662230
H&M Hennes & Mauritz AB	Apparel and Accessory Stores	56	HNNMY.ST	SE0000106270
Holmen AB	Paper and Allied Products	26	HOLM.ST	SE0011090018
Svenska Handelsbanken AB	Depository Institutions	60	SHBA.ST	SE0007100599
Swedbank AB	Depository Institutions	60	SWEDA.ST	SE0000242455
Skanska AB	Engineering and Management Services	87	SKA-B.ST	SE0000113250
Ratos AB	Security and Commodity Brokers	62	RATO.ST	SE0000111940
Securitas AB	Business Services	73	SECU-B.ST	SE0000163594
Industrivärden AB	Holding and other Investment Offices	67	INDU-A.ST	SE0000190126
Wihlborgs Fastigheter AB	Real Estate	65	WIHL.ST	SE0011205194
SSAB AB	Primary Metal	33	SSABA.ST	SE0000171100
Swedish Match AB	Tobacco Products	21	SWMA.ST	SE0015812219
Trelleborg AB	Chemicals and Allied Products	28	TRELB.ST	SE0000114837
Telefonaktiebolaget LM Ericsson	Electronic and Other Electric Equipment	36	ERIC.ST	SE0000108656
Boliden AB	Metal Mining	10	BOL.ST	SE0015811559
Alfa Laval AB	Industrial Machinery and Equipment	35	ALFA.ST	SE0000695876
Skandinaviska Enskilda Banken AB	Depository Institutions	60	SEBA.ST	SE0000148884

AB SKF	Instruments and Related Products	38	SKF.ST	SE0000108227
Telia Company AB	Communication	48	TELIA.ST	SE0000667925
Lundin Energy AB	Oil and Gas Extraction	13	LUNE.ST	SE0000825820
Electrolux AB	Wholesale Trade Durable Goods	50	ELUX-B.ST	SE0016589188
Svenska Cellulosa SCA AB	Paper and Allied Products	26	SCA-B.ST	SE0000112724
Sandvik AB	Primary Metal	33	SAND:ST	SE0000667891
Volvo AB	Engineering and Management Services	87	VOLV-B.ST	SE0000115446
Tele2 AB	Communication	48	TEL2:ST	SE0005190238
JM AB	Real Estate	65	JM:ST	SE0000806994
L E Lundbergföretagen AB (publ)	Real Estate	65	LUND-B:ST	SE0000108847
NCC AB	Construction Special Trade Contractors	17	NCC:ST	SE0000117970

* SIC: Standard Industrial Classification

Table 12: Switzerland Sample

This table reports industry names, SIC codes, tickers and ISIN codes for all Swiss firms included in common sample size 1 with corresponding ratings across Bloomberg and Refinitiv between 2012-2020.

Sample Statistics	(1)	(2)	(3)	(4)
Firm Name	Industry Name	SIC	Ticker	ISIN Code
Vifor Pharma AG	Chemicals and Allied Products	28	VIFN:SW	CH0364749348
Helvetia Holding AG	Insurance Carriers	63	HELN:SW	CH0466642201
Basilea Pharmaceutica AG	Chemicals and Allied Products	28	BSLN:SW	CH0011432447
Chocoladefabriken Lindt & Spruengli AG	Food and Kindred Products	20	LISN:SW	CH0010570759
Valora Holding AG	Food Stores	54	VALN:SW	CH0002088976
EFG International AG	Security and Commodity Brokers, etc.	62	EFGN:SW	CH0022268228
Georg Fischer AG	Rubber and Miscellaneous Plastics	30	GF:SW	CH0001752309
Partners Group Holding AG	Security and Commodity Brokers, etc.	62	PGHN:SW	CH0024608827
Swiss Prime Site AG	General Merchandise Stores	53	SPSN:SW	CH0008038389
PSP Swiss Property AG	Holding and other Investment Offices	67	PSPN:SW	CH0018294154
Straumann Holding AG	Holding and other Investment Offices	67	STMN:SW	CH0012280076
Sulzer AG	Holding and other Investment Offices	67	SUN:SW	CH0038388911
Rieter Holding AG	Holding and other Investment Offices	67	RIEN:SW	CH0003671440
Ems Chemie Holding AG	Chemicals and Allied Products	28	EMSN:SW	CH0016440353
OC Oerlikon Corporation AG Pfaeffikon	Holding and other Investment Offices	67	OERL:SW	CH0000816824
Temenos AG	Business Services	73	TEMN:SW	CH0012453913
Meyer Burger Technology AG	Electronic Equipment	36	MBTN:SW	CH0108503795
Belimo Holding AG	Industrial Machinery and Equipment	35	BEAN:SW	CH1101098163
Emmi AG	Lumber and Wood Products	24	EMMN:SW	CH0012829898
Huber+Suhner AG	Primary Metal	33	HUBN:SW	CH0030380734
GAM Holding AG	Holding and other Investment Offices	67	GAM:SW	CH0102659627
Banque Cantonale Vaudoise	Depository Institutions	60	BCVN:SW	CH0531751755
Dormakaba Holding AG	Business Services	73	KABN:SW	CH0011795959
Allreal Holding AG	Automotive Dealers and Gasoline Stations	55	ALLN:SW	CH0008837566
Mobimo Holding AG	Real Estate	65	MOBN:SW	CH0011108872
Flughafen Zuerich AG	Transportation By Air	45	FHZN:SW	CH0319416936

* SIC: Standard Industrial Classification

Table 13: Quintile Rankings Sweden (BB & RE)

These tables report the firms within the common sample which are consistently included in the 20% best- or worst-in-class firms for Bloomberg and Refinitiv. Panel- A and B show the best- or worst-in-class Swedish firms (tickers) before and after the implementation of the NFRD. See Table 11 for firm names.

Panel A: Quintile Rankings Before the Non-Financial Reporting Directive

	2012		2013		2014		2015		2016	
	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile
Total*	ATCO	HEXA	ELUX	SECU	SCA	SECU	BILL	RATO	SCA	SECU
	SCA	INVEB	SCA	HEXA	ELUX	INVEB	SCA	SECU	BILL	RATO
	BOL	SECU		RATO	SEBA	HEXA	BOL	GETI	BOL	INVEB
		GETI		GETI		GETI		HEXA		HEXA
		RATO		INDU		RATO		INVEB		INDU
		INDU		LUND		LUND		LUND		LUND
		LUND				INDU		INDU		
E**	SEBA	SECU	SCA	RATO	SCA	GETI	SCA	RATO	SKF	NCC
	BOL	HEXA	BOL	SECU	BOL	SECU	BOL	GETI	BILL	GETI
		LUNE		HEXA		HEXA		LUNE		HEXA
				LUNE		LUNE		SECU		SECU
								HEXA		
S***	MTGB	JM	MTGB	JM:ST	ATCO	GETI	MTGB	INVEB	SCA	HEXA
	ATCO	GETI	BILL	GETI	MTGB	RATO	SCA	RATO	MTGB	INVEB
	SWMA	RATO		RATO	SCA	INDU		INDU		RATO
		INDU		INDU						INDU
G*	ERIC	INDU	ATCO	LUND	CAST	INDU		AXFO	TELIA	LUND
	ATCO	LUND	BILL	SKA	SEBA	SKA		SKA	SCA	INDU
		SKA	TELIA	GETI	TELIA	GETI		SSABA		SKA
				INDU				INDU		

Panel B: Quintile Rankings After the Non-Financial Reporting Directive

	2017		2018		2019		2020	
	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile
Total*	SCA	SECU	BILL	RATO	ELUX	SECU	ELUX	SECU
	BILL	NCC		INVEB	BOL	INVEB	SCA	INVEB
		RATO		HEXA		HEXA	BOL	HEXA
		INVEB		INDU		INDU		INDU
		HEXA		LUND		LUND		LUND
		INDU						
		LUND						
E**	SKF	GETI	SKF	RATO	SKF	RATO	SCA	HEXA
		HEXA		HEXA		HEXA		SECU
		SECU		SECU		SECU		
S***	MTGB	HEXA	SCA	INVEB	NOBI	GETI	SCA	HEXA
		NCC:STO		HEXA	ALFA	HEXA	BOL	
		RATO						
		INDU						
G*	SAS	SKA	LUNE	WIHL	LUNE	LUND	LUNE	ASSA
		LUND		LUND		HEXA		LUND
		INDU		HEXA				WIHL
								INDU
								HEXA

* With 41 firms included in the sample size, 8 firms are included in the top- and bottom quintiles (20%)
** With 37 firms included in the sample size, 7 firms are included in the top- and bottom quintiles (20%)
*** With 39 firms included in the sample size, 7 firms are included in the top- and bottom quintiles (20%)

Table 14: Quintile Rankings Switzerland (BB & RE)

These tables report the firms within the common sample which are consistently included in the 20% best- or worst-in-class firms for Bloomberg and Refinitiv. Panel- A and B show the best- or worst-in-class Swiss firms (tickers) before and after the implementation of the NFRD. See Table 12 for firm names.

Panel A: Quintile Rankings Before the Non-Financial Reporting Directive

	2012		2013		2014		2015		2016	
	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile
Total*	SUN	SPSN	SUN	SPSN	SUN	ALLN	GF	ALLN	GF	ALLN
	GF		GF		GF	TEMN	SUN	EMSN	SUN	EFGN
	STMN		STMN		STMN	EMSN	STMN	EFGN	MBTN	
									KABN	
E**	GF	GAM	SUN	GAM	SUN	BCVN	SUN	BCVN		BCVN
	SUN		GF	BCVN						
S***	GF	PGHN	GF	PGHN	SUN	EFGN	GF	EFGN	GF	PGHN
	SUN	EFGN	SUN	EFGN	GF	PGHN	STMN	PGHN	STMN	
	STMN				STMN				RIEN	
G*	SUN	FHZN	SUN	MOBN	GF	FHZN	GF	FHZN	MBTN	SPSN
	GF	LISN	GF	FHZN	SUN		SUN		GF	FHZN
			STMN				MBTN			

Panel B: Quintile Rankings After the Non-Financial Reporting Directive

	2017		2018		2019		2020	
	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile
Total*	GF	BCVN	GF	EMSN	GF	BSLN	KABN	EMSN
	MBTN	EMSN	KABN	EFGN	KABN	EMSN	GF	EFGN
					VIFN	EFGN		
E**	KABN	GAM	KABN	GAM	KABN	BCVN	KABN	BCVN
		BCVN		BCVN			GF	
S***	GF	FHZN	GF	EFGN	GF	FHZN	GF	BSLN
	RIEN	EFGN	KABN		RIEN	EFGN	RIEN	FHZN
								EFGN
G*		FHZN	GF	FHZN	HELN	LISN		FHZN
						EMSN		EMSN
						FHZN		

* With 26 firms included in the sample size, 5 firms are included in the top- and bottom quintiles (20%)

** With 17 firms included in the sample size, 3 firms are included in the top- and bottom quintiles (20%)

*** With 19 firms included in the sample size, 3 firms are included in the top- and bottom quintiles (20%)

Table 15: Demographic and Economic Comparison Switzerland and Sweden

Characteristics	Switzerland	Sweden
Population	8.5 million	10.3 million
Labor force	5.067 million	5.029 million
Labor force by occupation	76.9% services, 19.8% industry, 3.3% agriculture	86% services, 12% industry, 2% agriculture
Real GDP - PPP (2017 $)	$590 billion	$525 billion
Credit rating	AAA	AAA
GDP composition by sector	73.7% services, 25.6% industry, 0.7% agriculture	65.4% services, 33% industry, 1.6% agriculture
GDP composition by end use	53.7 % household consumption, 12 % government consumption, 24.5% investment in fixed capital, -1.4% investment in inventories, 65.1% exports of goods and services, -54% imports of goods and services	44.1% household consumption, 26% government consumption, 24.9% investment in fixed capital, 0.8% investment in inventories, 45.3% exports of goods and services, -41.1% imports of goods and services
Agricultural products	Milk, sugar beet, wheat, potatoes, pork, barley, apples, maize, beef, grapes	Wheat, milk, sugar beet, barley, potatoes, oats, rapeseed, pork, rye, triticale
Industries	Machinery, chemicals, watches, textiles, precision instruments, tourism, banking, insurance, pharmaceuticals	Iron and steel, precision equipment (bearings, radio and telephone parts, armaments), wood pulp and paper products, processed foods, motor vehicles

Data retrieved from CIA World Factbook data

Table 16: Definition of Variables

This table reports the dependent and independent variables that are used in our empirical model.

Variables	Definition
Dependent	
Total Score Correlation (TSC)	The correlation for the common sample for Refinitiv Total Score and Bloomberg Total Score
Environmental Score Correlation (ESC)	The correlation for the common sample for Refinitiv Environmental Score and Bloomberg Environmental Score
Social Score Correlation (SSC)	The correlation for the common sample for Refinitiv Social Score and Bloomberg Social Score
Governance Score Correlation (GSC)	The correlation for the common sample for Refinitiv Governance Score and Bloomberg Governmental Score
Independent	
Treatment Dummy (Treat)	Treatment dummy: Sweden = 1, Switzerland = 0
Post Regulation Dummy (Post)	Post Regulation Dummy: 1 >2016
Regulation Dummy (TREAT x POST)	POST x TREAT dummy: 1 = (TREAT and POST) = 1