ESG Score Prediction & Public Awareness

Student Name & ID: Winston Arendsen - 453480

Thesis Supervisor: Gangaram-Panday, Y.S.

Second Assessor: Dr. Lemmen, J.J.G.

Abstract

The debate on the role and responsibilities of companies in our society becomes increasingly important in light of current developments such as global warming. One aspect where this debate materializes is in the rise of sustainable finance where investors try to, at the very least, persuade companies to do no harm to the environment with their business. However, measuring the level of corporate social responsibility (CSR) of a firm is a challenging task. One metric that attempts to measure the level of CSR is the Environment, Social and Governance (ESG) score. ESG scores are increasingly important in investment decisions making and are an attempt to measure the corporate social performance of firms. However, these scores are time consuming and costly to produce, and subsequently not widely available for many firms. This research tries to add to the literature that overcomes the aforementioned difficulties with ESG scores by predicting these scores based on publicly available information. In particular, this research attempts to measure the influence of the public perception and awareness on ESG score by using advertising intensity and the reputational risk index. Using random and fixed effects estimation techniques and data from North American companies this research found a significant effect of public awareness and perception on the ESG score. Moreover, using financial performance measures resulted in mixed results. Some financial performance measurements have a significant effect on ESG scores but most used in this research do not.

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Financial Economics

April 30, 2022

This page is left blank intentionally

Contents

Primum non nocere.

First, do not harm. One of the principles of sustainable finance.

1 Introduction

In most of finance and corporate finance the goal of any firm is maximize its value and thus, maximize value for shareholders. To do so, a firm should focus on profit maximization to assure the most efficient use of the capital provided by its shareholders. In this context, a manager of a firm that spends money on other activities than those who generate profit is not desirable because it decreases the efficiency of the capital allocated to the firm.

In an article in the New York Times in 1970 Milton Friedman argued that the social responsibility of a business is to make profit. He argues that a corporation cannot have a responsibility, only people can. According to him business executives are mere employees of the stockholders and should not invest in other activities in charities or other socially acceptable manners since it is not their own money they are investing with.

Such an economic goal however is only focused on one stakeholder of the firm, the shareholder. Consequently, this approach can lead to external effects, in particular negative external effects. Externalities arise when property rights are unclear or nonexistent Harvey S. Rosen (2014). To put it more simply, firms can make costs that they do not have to pay for. One prominent example of such an externality is the air-pollution externality. Whereby a firm pollutes the air by emitting greenhouse gases such as CO2 into the air in the process of providing a service or product. This can have a serious impact the on climate and this can lead to costs for society, for example, when extreme weather events take place and destroy property. The problem is of course that with the complexity and "long-term" nature of the climate it is hard to point to any entity for its direct involvement in making these cost. Rather, cost can come infrequent and over a longer period of time. Since no entity is owner of all the air in the world anyone, including firms, can pollute the air with little to no impact on their individual welfare and thus giving rise to an externality since it does come at a cost for all users of air. As stated in a 2021 report of the World Economic Forum on global risk, failure of climate change mitigation and adaptation was one of the biggest risks of our time, with a potential impact bigger than weapons of mass destruction *The Global Risks Report 2021* (2021). However, negative externalities with regard to pollution are not the only externalities firms can produce. There are all sorts of negative externalities firms can produce.

Externalities are not just bad for entities directly involved in the matter but can have serious economic consequences. This is because externalities lead to imperfect markets and thus inefficient equilibrium's, in other words, there is an inefficient use of resources Gruber (2013).

To tackle these negative externalities and to get a better insight in the social impact and performance of businesses academics and practitioners try to measure beyond the monetary impact of a firm but also the broader economic and social impact. Looking beyond monetary efficiency is a form of corporate social responsibility (CSR) and is a topic that not only looks at economic values but also ethics and public perception on what the role of businesses is and should be in society. Measuring beyond basic economic values however is a complex topic. An entire branch of research is dedicated to modelling predicting and measuring this this corporate social performance. Important models include those of Wood (1991) and Mitnick (2000) who systematically approach the topic.

One central problem to the topic of corporate social performance (CSP) and CSR measurement is the inherent lack of quantitative and qualitative data. This problem is partly induced by the complexity of some of the topics that need to be measured. Some data vendors started to keep track of social performance of corporations using so called ESG scores, which stands for environment, social & governance. Socially responsible investments are not just a niche but are developing into an investment approach with significant material impact as the ESG as can be seen from its rapid growth in assets under management from 22.8 trillion US dollars in 2016 to 37.9 trillion US dollars by the end of 2021 and expected to keep growing for the coming years [1].

One drawback of current ESG-data is that its time consuming and costly to thoroughly assess the performance of a company with respect to the ESG factors and drivers. This is because all raw data on the separate ESG topics needs to be collected, for all individual companies, analyzed and processed into an ESG score. In their report on ESG investing the OECD compares the methodology of the most commonly used ESG ratings Boffo & Patalano (2020). These methodologies use a complicated process of selecting, collecting, judging different ESG factors and indicators and translating that information to ESG scores for each individual firms. For example, the Refinitiv ESG scores uses between 70 and 170 datapoint per ESG score depending on the industry [2].

Some researchers and practitioners have therefore tried to investigate the relationship between ESG scores and publicly available variables to test to what extent ESG scores can be predicted based on other factors (Garcia et al. (2020), Krappel et al. (2021), D'Amato et al. (2021), Licari

[1] https://www.bloomberg.com/professional/blog/esg-assets-may-hit-53-trillion-by-2025-a-third-of-global-aum/

[2] See: https://www.refinitiv.com/content/dam/marketing/en$_us$/*documents/methodology/refinitiv − esg − scores − methodology.pdf*

et al. (2021)). This is, however, a relatively new branch of research with the only 4 papers written on this topics are all from 2020 and 2021. Not all relationships have been researched so far. One interesting relationship that has not been further analyzed based on the theoretical models of CSP measuring is the relationship between ESG scores and public awareness and perception. Moreover, different techniques and data sets have also led to mixed results on the effect of different factors on CSP factors.

This research aims to further develop insights on the complex relationships of ESG scores, corporate social performance, corporate financial performance, public awareness and public perception to enhance the predictive ability of ESG predicting models. The following research question will be central in this research:

To what extent do CSP outcomes and public perceptions drive ESG scores of firms?

To clarify, CSP outcomes are the measurable outcomes of the CSP processes. For example, the CO2 emission of a company, or the diversity of the board. Note that this topic will be further explained in the literature section of this study. Although ESG scores are a summary measurement of the CSP outcomes, they are not simply an arithmetic addition of all CSP outcomes but rather a comprehensive way to view the overall performance of a firm relative to its peers. Hence, some CSP outcomes will have a greater impact on the ESG score of a company than others. Moreover, many CSP outcomes in raw format are a far cry from the quantitative nature of ESG scores. For example, a CSP outcome might be that a company decides to lower the CO2 emission. It is, however, not immediately clear what impact this might have on the overall corporate social performance, or how socially responsible the firm is after lowering its CO2 emission. ESG scores, in essence, try to quantify this kind of information.

The rest of this paper will be structured as following. First, a literature review will be performed to find the most important theoretical and empirical literature on corporate social responsibility (CSR), corporate social performance (CSP) and environment, social and governance (ESG) scores and ESG predictions. Next, the knowledge of the literature review will be summarized in a conceptual model for ESG score prediction. After that three hypothesis will be formulated based and the research outline is formed. Next, the data and methodology used to test the hypothesis will be discussed. After this the results will be interpreted, discussed and a conclusion will be drawn based on the results and subsequent discussion. Finally, the limitations of this paper will be discussed and suggestions for further research are presented.

2 Literature Review

To better understand ESG, ESG scores, and ESG score predictions it is useful to first explore where it comes from and what the current and past literature has already written on the topic. This literature review will start by reviewing where corporate social responsibility comes from and how it has developed over time. Next, it will further explore a subtopic of CSR which is the topic of corporate social performance and measurement. It will then review literature on ESG and ESG predictions. Next, a conceptual model for ESG prediction based on the literature reviewed will be constructed. Finally, hypothesis and sub-questions will be formulated based on the literature review and conceptual model.

2.1 Corporate Social Responsibility (CSR)

One possible way to decrease firm induced externalities via can be via corporate social responsibility (CSR), a way for firms to self-regulate and to take into account multiple stakeholders instead of just the shareholder. This all to reduce the costs that incur whilst striving for their economic targets Weber & Wasieleski (2018).

Corporate social responsibility has a long history, and as with many complex definitions, it is hard to point out when and where exactly it originated from. Some of the first corporation like entities already existed since the Romans, where corporations were entities with mostly social purposes Chaffee (2017). Carroll (2008) argues in his book that the CSR trend we know today stems from the 1950s. When Howard R. Bowen published the book "Social Responsibilities of the Businessman" in 1953. In this book Bowen asks a very relevant question that is still under the CSR debate today:

> "What responsibilities to society may businessmen reasonably be expected to assume?"
> *- Bowen (1953)*

Today, CSR has developed into a more complex quest for firms to strive beyond profit. It is referred here as a complex quest because there are a lot of ambiguities surrounding the CSR topic. For instance, there is no clear consensus on the exact definition of CSR Sheehy (2015). This is partly due to the inherent nature of the concept, what is socially responsible today might be seen as a crime in the future and vice versa. Moreover, the topic is still being researched both from a practitioner's view as a theoretic. Carroll (1999) reviews the development of the concept of CSR through the decades starting in the 1950s.

One clear sign that firms are moving towards or trying to move towards a more stakeholder-oriented view instead of just shareholder was the declaration of 181 CEOs of large American companies to commit to lead their companies to benefit all stakeholders of a firm, a large shift from the earlier focus on shareholders Business Roundtable (2019). It should be noted here that this of course is a sign and not clear evidence of firms increasing their CSR effort. This new focus of firms is met some critique however. For example, Damodaran (2019) argues shifting the focus away from shareholders and seek benefit for all stakeholders of a firm can lead to conflicting interests and a decrease in the accountability of managers.

As in line with the argument of Milton Friedman, some of these new objectives for firms can be in direct conflict of interest with each other. For example, an increase in price to pay higher wages to employees might be a good social move, but if consumers then move to competitors with a lower price this can hurt the profit objective of the firm.

Some researchers have found evidence that firms value is positively related to firms that engage in CSR activities (Hu et al. (2018), Servaes & Tamayo (2013)).

Barnett (2007) argues that the reason there is still mixed results on the literature on the relationship between corporate financial performance (CFP) and CSR is because most researcher do not include that path dependency of the relationship firms have with their stakeholders. In other words, he argues that firms can have different returns on CSR investments even if they are in many aspects similar due to the different relationship they have with their respective stakeholders. Barnett's arguments are mainly based on the stakeholder theory partly derived from the work of Freeman (2010) where the purpose and impact of a business should not just be to create value for stockholders but for all stakeholders of the business.

Servaes & Tamayo (2013) further build upon the work of Barnett and looked whether customer awareness played a role in the relationship between firm value and CSR. Here customer awareness is measured as advertising intensity which is measured as advertising expenditure as a percentage of sale. Their results indicate that there is a positive relationship between CSR and firm value for firms with high customer awareness. This relationship does not seem to hold or even inverts for firms with medium or low customer awareness. They conclude that CSR can add value to a firm but only under certain conditions.

Hu et al. (2018) further researched whether the positive relationship between CSR and firm value still holds outside of the standard researched countries which are predominantly North American and Western European countries. They focused on China and found again a positive relationship

between CSR and firm value using a linear regression model. However, Crisóstomo et al. (2011) found a negative relationship between firm value and CSR suggesting that CSR activity has a negative impact on the firm value in Brazil. The difference in their outcomes might be explained in the data used by both studies which depending can vary strongly on how and when you measure it. Which currently is an inherent problem with CSR research in general. For example, in the Crisóstomo et al. (2011) state that Brazilian firms are not obligated to disclose information about social actions, as is the case in most of the world, and are free in deciding which CSR activities are disclosed and which are not. Next, in their literature review, Ali et al. (2017) look into the determinants of CSR disclosure in developed and developing countries. Their analysis of 76 empirical research papers reveals that in developed and developing countries CSR disclosure is mostly driven by company characteristics (e.g., company size, governance mechanism etc.). An important difference however in the drivers of CSR disclosure between developing and developed countries is in their respective stakeholders. In developing countries stakeholders are predominantly external (foreign) (e.g., international media or foreign investors). Whilst in developed countries concern of specific stakeholders and the public for CSR disclosure, thus stakeholders are more internal (more national).

Next, Jo & Harjoto (2011) investigate if the choice of corporate governance, both internal and external, has an effect on the choice of CSR engagement and proceed to investigate whether engagement in CSR activities influences firm value. They found a positive relationship between firm value and CSR engagement. Moreover, their results show that external monitoring over a firm's CSR engagement by a securities analyst has a more significant effect on firm value than other internal and external governance mechanisms.

2.2 Corporate Social Performance (CSP)

As previously mentioned, a problem when doing CSR related research is that it can be hard to measure the CSR performance of a firm. In essence a lot of CSR related topics are of qualitative nature and can be subjective depending on the stakeholder, i.e. corporate culture, diversity, employee compensation, sustainability etc. What one person might perceive as favorable for society someone else might view as destructive. Wood (1991) developed a model for Corporate Social Performance (CSP). CSP is a concept related to CSR. In her model Wood (1991) tries to bridge the gap between CSR and social responsiveness, that is to what extent is a firm willing and able to respond to social pressures. Wood (1991) builds upon the work and definitions of Wartick & Cochran (1985), who

developed a three-facet model for CSP, to come to an own definition of CSP:

> "A business organization's configuration of principles of social responsibility, processes of social responsiveness, and policies, programs, and observable outcomes as they relate to the firm's societal relationships" Wood (1991)

It is clear from his definition of CSP that to assess a firm's social performance one should examine three facets which he defines in his model. Table 1 shows Wood's model.

The Corporate Social Performance Model
Principles of corporate social responsibility
Institutional principle: legitimacy
Organizational principle: public responsibility
Individual principle: managerial discretion
Processes of corporate social responsiveness
Environmental assessment
Stakeholder management
Issues management
Outcomes of corporate behavior
Social impacts
Social programs
Social policies

Table 1: Source: Wood (1991) p.694

This model helps to more clearly define how to and what to measure when looking at corporate social performance. The model shows that CSP is thus not just about the outcomes, one needs to define what needs to be measured and how to measure it before the outcomes can be properly analyzed. Hence, one needs to analyze the three facets of CSP as identified by the model of Wood (1991).

First, one needs to examine the principles of corporate social responsibility. Although the fundamental idea behind CSR is that business and society are strongly intertwined and thus not separate entities Wood (1991) argues that there are three clearly different conceptual approaches to identify the principles of corporate social responsibility. Namely,

- **Legitimacy**, Wood (1991) derives this principle from the work Davis (1973). In his article Davis presents for and against business assumption of social responsibility. One argument for CSR states that a business gets its legitimacy and power from the society. Companies that do not use their power according to what society considers responsible tend to lose their power. However, Davis also notes that process of taking back power can be a long-run process taking decades or even centuries.

- **Public responsibility**, Wood (1991) derives this from the work of Post & Preston (2012) first published in 1975. This principle states that the social responsibility of a firm only goes to primary and secondary areas a business operates in. To put it simply, a small business that sells clothing online might be responsible for good working condition, but not for building a dam to protect people from floods on the other side of the country.

- **Managerial Discretion**, the principle of managerial discretion in Wood's model states that a manager is a moral actor within the business organization and are obligated to act upon CSR related social topics with discretion to create outcomes that are socially responsible. Wood (1991) bases this largely on the works of Corroll (1979) and Wood (1990).

Second, one needs to examine the social responsiveness of a firm. Wood (1991) includes social responsiveness as one of the three facets rather than the only facet based on the argument of Corroll (1979), who argues that a firm might be willing and able to respond to social pressure, but whilst doing so can slack on their true responsibility or act irresponsibly in trying to act on another responsibility. The social responsiveness part of the CSP model compliments the CSR component by explaining how to translate the normative concepts established by CSR to actionable responses. Wood (1991) considers three characteristic behaviors of responsive firms which he in part derives from the work of Ackerman (2013) first published in 1975; environment assessment, stakeholder management and issues management. Environment assessment states that a business's survival is dependent on its adaption to the environment. It is thus important for a firm to analyze the environment it operates in. In the CSP model it establishes the context of a social issue to which a business might need to respond to. Next, stakeholder management analyses the relationship between a firm and its relevant stakeholders. It thus provides an analysis of the relevant actors to a social issue a firm might need to respond to. Finally, the third part of the social responsiveness facet of the CSP model is issues management (IM). IM according to Wood (1991) should be interpreted as the process of social responsiveness by an organization. Wood is very clear that this should not be seen as the endpoint of CSP as Wartick & Cochran (1985) conclude in their work. IM can thus be seen as the analysis of interests when a firm is faced with a social issue.

The third and last facet of the Wood (1991) CSP model are the observable outcomes of corporate behavior. Wood (1991) divides the outcomes into three subcategories: social impacts, social programs and social policies. The social outcomes of any decisions a firm makes are the only observable outcomes of the CSP model motivations, principles etc. are not clearly observable. Wood

concludes that more research needs to be done on the CSP model specifically in the outcomes part of the model.

It should be noted that Wood revised her model later with the central idea and structure remaining but changing the outcomes facet of the model (Wood (2010).

2.2.1 Measuring and Predicting Corporate Social Performance

The model of Wood (1991) described above helps to understand where CSP comes from and what can and cannot be measured. However, measuring CSP can be a difficult task due to the complex and at times subjective nature of the topics making up CSP. Mitnick (2000) looks into the issues surrounding CSP measurement which he sees as part of the literature on the social audit. Mitnick (2000) does this by introducing three metrics to measure CSP as such to establish a basis for a systematic theory on CSP measurement. The metrics he introduces are: the metric of performance evaluation, the metric of performance measurement and, the metric of performance perception and believe. In other words, to measure CSP one needs to know first what we want (what we value). Next, one need to know (measure) what is done to get what we want, and finally, we need to measure if we believe what is claimed on the progress of getting to what we want is true. In his paper Mitnick (2000) focuses on the latter two metrics because these focus on the outputs and outcomes of his model which are more straightforward to measure than the input of the model, what we value and how much we value it.

To fully understand the systematic measuring of CSP proposed by Mitnick (2000) a short summarize the three metrics proposed in his paper.

- Metric of performance evaluation: tries to measure what we value and how much we value it. Although one can measure what choices a person can make to reach a Pareto optimal situation, that is a situation where any trade-off results in a situation where no one can improve their situation without making someone else worse off. Such an exercise would not measure why a person values something.

- Metric of performance measurement: this measures the achievements of CSP which in itself are defined by the values measured with the metrics of performance evaluation. Mitnick (2000) starts by arguing that academics should not necessarily aim to create models of CSP measurement that act as a simplification of reality. He argues that a lot of academics use the Ockham's Razor in a misguided manner. When something is inherently complex, such as CSP

measurement, the theories describing it should be able to be complex as well. Next Mitnick (2000) starts developing a theoretical framework on systematic CSP measurement. His simple system consists of where inputs (i.e. values) are converted under guidance of a system of governance to produce output activities which results in outcomes that can provide feedback to the inputs of the system. This system is to some extents somewhat comparable to the CSP model of Wood (1991) described earlier in this paper. Mitnick (2000) further develops the system by introducing six system performance domains of which performance measurements can be created. In figure 1 an overview of these system performance domains is given, and some common measurements used in each domain.

- Metric of performance perception and belief: is defined by Mitnick (2000) as the measurement to what extent the public beliefs and perceives the claims of achievements by businesses (or social institutions) measured. Mitnick (2000) includes this as a third measurement since what the public perceives and believes can be different than what measurement tells you. He argues that a firm might commit to a certain value (the first metric) and subsequently has valid and reliable results on the performance regarding this value (the second metric) but once the firm claims these results key stakeholder might refuse to acknowledge these claims. Following this argument, it does not matter whether you have valid instruments for CSP measurements if the public does not believe the results. Based on his own developed theory of testaments Mitnick (2003) argues that the following will generally hold true:

  Reports are preferred to claims, and claims are preferred to predictions.

Mitnick (2000) extensively discusses a number of propositions on the last two metrics of is model which he believes are generally true to when attempting to measure either the CSP performance or the perception or believes of the public on CSP reports, claims and predictions. His model helps to identify what, and how to measure the complex topic of CSP. Moreover, his work also helps to explain some key characteristics of CSP measurements.

With this extensive theoretical foundation for CSR, CSP, and CSP measurements it is interesting to look what topics have already been researched. Wood (2010) argues that is an over emphasis by CSP research in some areas whilst other areas deserve some more attention. This is partly due to an inherent lack of good and accessible data on CSP. This lack of data has many reasons. However, many businesses have a good reason to not share too much data on their activities and impact of their activities on their respective environment (e.g. information that given them a competitive

Table 2
System Performance Domains and CSP Measures

SYSTEM PERFORMANCE DOMAINS:	*Guidance/ Norm-setting*	*Consumptive/ Contributory*	*Process*	*Activity*	*Outcomes*	*Summative/ Combinatorial*
CSP MEASURES:	Analysis of mission statements	Measures of hiring goals	Measures of diversity in workforce	Social audits of activities, including philanthropy	Measures of the political knowledge of employees	Reputational measures: KLD, *Fortune,* and so on
	Codes of conduct for various company activity areas	Information disclosure in hiring practices	Measures of corporate misconduct such as white-collar crime	Percentage of pre-tax earnings to charity	Measures of an unpolluted environment	
	Subscription to external guides or sets of principles:	Measures of purchasing restrictions	Degree of cooperation with regulators	Toxic Release Inventory (TRI)	Measures of a good economic environment	
	Adherence to Sullivan principles		Reporting on regulatory fines	Social disclosure in annual report	Civil cases brought against the corporation	
	Adherence to Valdez principles					

Note: CSP = corporate social performance; KLD = Kinder, Lydenberg, Domini & Co. measures.

431

Figure 1: Source: Mitnick (2000) p.431

advantage),

One-way investors and other stakeholders can get more information about the CSP of a firm is with social ratings and scores. These ratings are an attempt to measure the social performance of a firm with respect to a variety of disciplines. Mostly these include some environmental performance, such as These are scores based on the assessment of rating agencies. Two databases that are currently popular are the ASSET4 database provided by Refinitiv and the MSCI ESG STATS which are formerly known as KLD stats.

Chatterji et al. (2009) are some of the first researchers to empirically research whether companies with good social ratings are actually a good transparent measure of past and future environmental performance. For they try to predict past and future environmental ratings based on their own collected environmental data on the companies with a social rating. Their results suggest that the KLD ratings on environmental performance provide a reasonably good measure of past environmental performance. In addition, the KLD environmental strength ratings do not seem to be a good predictor of future environmental performance.

2.2.2 CSP Data and Validity

Due to mixed results in the literature on environmental rating validity Semenova & Hassel (2015) have looked into the comparability of the different ratings provided by different rating agencies

that are available to researchers and practitioners these days. The results indicate that although the different ratings have different commonalities, on aggregate, they do not converge. This is in line with literature on environmental performance that says it's difficult to measure environmental performance. In addition, it confirms one of the fundamentals CSP measurements (Mitnick, 2000) where corporate social performance is inherently complex and thus its measurements can be complex. Hart & Sharfman (2015) also looked at the validity of the KLD (MSCI ESG STATS) data and found that after a change in measurements in the KLD database itself, the validity is concurrent with the original KLD database. Moreover, there results also indicated that continues measurements perform better than binary measurements. Again, in line with what is expected from previous literature ons CSR and CSP. Lastly, Chatterji et al. (2016) also found evidence for strong evidence for low commensurability between six largest socially responsible index (SRI) rating vendors. These results further underscore the importance to be careful with interpreting the results form CSP related research until.

2.2.3 The Business Case for CSP

One aspect that got a lot of attentions from researchers and practitioners is the "business case" for CSR. This stems from the idea that businesses might profit from doing good to society in a multitude of ways. Some include more loyal customers, some sustainability of businesses or a lower idiosyncratic risk. The relationship between CSP and corporate financial performance (CFP), or in short financial performance (FP), (Wood, 2010).

The results on the relationship between a firm's FP and CSP however, have some mixed results in the literature. Ruf et al. (2001) found a positive relationship a firm's FP and CSP activity, specifically a positive relationship between CSP and sales. Moreover, Waddock & Graves (1997) found a positive relationship between CSP and a firm's FP. However, the direction they found is between FP and CSP. Meaning that a firm that has a good FP also is more likely to have good CSP. Which intuitively makes sense since a firm first wants and needs to have good financial results before it incorporates other stakeholders in its processes. Wang et al. (2016) performed a meta-analysis of 42 studies on the CSR and FP relationship. He found a significant positive relationship between CSR activity the FP of firms.

However, other researchers were not able to find a significantly positive relationship between CSP and FP.Makni et al. (2009) found no significant relationship between a firm's FP and CSP performance when looking at Canadian firms. Soana (2011) also found no significant relationship

between CSP and FP in the banking sector using ethical ratings provided by Ethibel and AXIA. Van de Velde et al. (2005) critique the literature CSR and FP for not including R&D expenses as a control variable when researching the relationship between CSR and the FP of firms. They found that when controlling for this variable the relationship between CSR and FP of firms becomes neutral instead of significant and positive.

Lastly, Barnett & Salomon (2006) found a mixed relationship between an investment fund's financial performance and its level of social screening. That is the practice where investors screen and select companies based on social and environmental criteria. Barnett & Salomon (2006) results show that when the level of screening increases the financial performance decreases but inverses when screening reaches a maximum level. They call this a curvliniear relationship. These results seem to be in line with the literature that argues that for more complex models to describe and measure social performance of firms.

It should be noted, as stated in prior research that one reason for this apparent discrepancy between the results might be in the data, and time period used for each paper. For example, Ruf et al. (2001) use data from the KLD database (MSCI ESG STATS) for the period 1992-1995 whilst Makni et al. (2009) use data from the Canadian Social Investment Database in the period 2004-2005. Moreover, Soana (2011) uses data provided by Ethibel and AXIA at one point in time (31/12/2005 and 31/12/2005 respectively).

This makes it more difficult to compare the different papers on CSP with each other and to make more generalized conclusions on CSP or even on CSP subtopics (e.g. environmental performance)

A stated earlier, this research is partly based on the premises that doing good might be profitable for a business. However, Wood (2010) argues that a firm's financial performance was only part of the overall social performance of a firm. She also argued that the search for a statistical relationship between a firm's financial performance and social performance can be misguiding. Another reason for this "over-emphasis" on the CSP-FP relationship is simply due to a lack of good reliable data CSP measures.

2.3 Environment, Social & Governance (ESG)

Data on CSR related topics can be scarce for a number of different reasons. One reason is that firms are not always obligated to publish data on their CSR performance. Moreover, it might not always be beneficial for a firm to publish too much information on their business process and outcomes. Another reason, data on CSP measures and CSR topics can vary from source to source and at times

can be inconsistent when multiple data sources report on the same topic. This is partly due to the complex nature of the topic. One way researchers and practitioners can overcome this issue is by using more standardized measurements for CSP. One of the more standardized methods are Environment, Social and Governance scores, also known as ESG scores or ESG ratings. However, even ESG-data vendors do not provide completely consistent data as is shown by Gibson Brandon et al. (2021) showing that there is a positive relationship between ESG disagreement and stock returns.

The terms ESG was first reported in 2005 at a conference for asset managers [3].

ESG-scores are produced by rating agencies and try to bridge the gap between the qualitative nature of the data and the need for quantitative data that can be used for analysis. This is, among other things, done by using standardized data points in each three of the ESG categories and calculating an average score based upon all these data points combined. The benefit of using ESG-scores is that it makes it easier to interpret the state of CSP/ESG-performance of companies and to compare them. However, a downside is that ESG-scores are a summary of the underlying measurements and data and thus contain less information than individual metrics.

One problem with ESG-data is that a lot of it is produced by commercial vendors.

This is a problem because this can lead to a bias in the companies selected and rated. And thus, a selection bias in the ESG-data. Moreover, due to the complex nature of the topic, rating companies based on a large set of ESG criteria can be a time-intensive and thus costly undertaking. This all leads to ESG data sets with mostly large, publicly listed companies. Drempetic et al. (2020) found a company size bias in the ASSET4 database, underlying the importance of carefully handling and analyzing ESG-data. Whilst the data on the companies that are in the ESG-databases might be very useful for analysis of these companies, to be able to draw more generalized conclusions or to better under the ESG topics more data and more diverse data is necessary. The importance of good data is once more underlined by Kotsantonis & Serafeim (2019).

2.3.1 ESG Prediction

To tackle this data collection problem researchers and practitioners are trying to predict ESG-scores using statistical models based on publicly available data. Since this is a relatively new branch of research only a few papers have been written on this subject.

[3] See: https://www.ifc.org/wps/wcm/connect/Topics_Ext_Content/IFC_External_Corporate_Site/Sustainability-At-IFC/Publications/Publications_Report_WhoCaresWins2005__WCI__1319576590784

Garcia et al. (2020) use a machine learning technique called the rough set approach to predict ESG scores based on firm financial variables. In essence a rough set approach is a machine learning technique that generates a set of decision rules based on the input variables (in this case financial variables) to group the predictions of the ESG scores. The maximum number of groups here is the number of firms in the data-set used, such a case would result in a "normal" regression. When the number of groups is reduced, so are exact ESG-score predictions. It can, however, increase the predictive power because the rough set algorithm identifies redundant or unimportant information in the data and thus resulting in better predictions as Garcia et al. (2020) show financial variables have predictive power over ESG scores, but this predictive power disappeared when larger groups were computed. Their model achieves an r-squared of about 27%, however, it does include some strongly significant variables. The optimal number of groups in their research is three or four equally balanced groups which reach a hit ratio of about 54-42%, which means that 52 to 42% of all predicted ESG scores fall in the correct group. This is a significantly better than the naive model where on average either 25 or 33% of all predictions are correct.

Krappel et al. (2021) also use machine learning techniques to predict ESG scores. They use ensemble models to predict ESG-scores based on publicly available fundamental data. Ensemble machine learning models are, as the name suggest, an ensemble of machine learning algorithms to enhance predictive ability. They are able to predict 54% of the variation (r-squared) in ESG scores. This is somewhat comparable to the 54-42% Hit ratio Garcia et al. (2020) got from their best rough set model. The fundamental difference is that Krappel et al. (2021) predict exact ESG-scores with this approach and Garcia et al. (2020) do not. They, however, do not use their own simple regression model as comparison but rather use the linear regression model computed by Garcia et al. (2020). Another important difference is that they incorporate some non-financial data this, however, is somewhat limited to other fundamental firm data, such as location of headquarter, industry etc.

Next, D'Amato et al. (2021) also use a machine learning technique to predict ESG scores. Their research uses balance sheet data of the companies with ESG scores. Using a random forest algorithm and comparing it to a generalized linear model (GLM) they found that balance sheet items have significant explanatory power on ESG scores.

Next to academics, practitioners also attempt to predict ESG scores.

Licari et al. (2021) from Moody's also developed a model based on ensemble machine learning techniques to predict ESG-scores and using a wide variety of variables on company, regional and country level. Just as Krappel et al. (2021) they use ensemble model building techniques, they

however, achieve an r-squared on their ensemble model of about 77%. Whilst their simple regression model achieves an r-squared of about 31%. The most important driver for ESG scores in their research was company size.

Interestingly all research up until this point on ESG prediction heavily rely on financial variables to predict ESG scores. Although they all found significant relationships as prior literature indicates that this can be expected due to for example data limitations, it also shows, as the researchers themselves also point out, it might give an incomplete results and conclusions.

2.4 Conceptual Model for ESG Prediction

Based on the literature review a conceptual model for ESG score prediction is constructed, figure 2 shows the conceptual model. Note that this is a simplified version of the theories described in the literature review to visually show how the main theories and ESG drivers relate to each other.

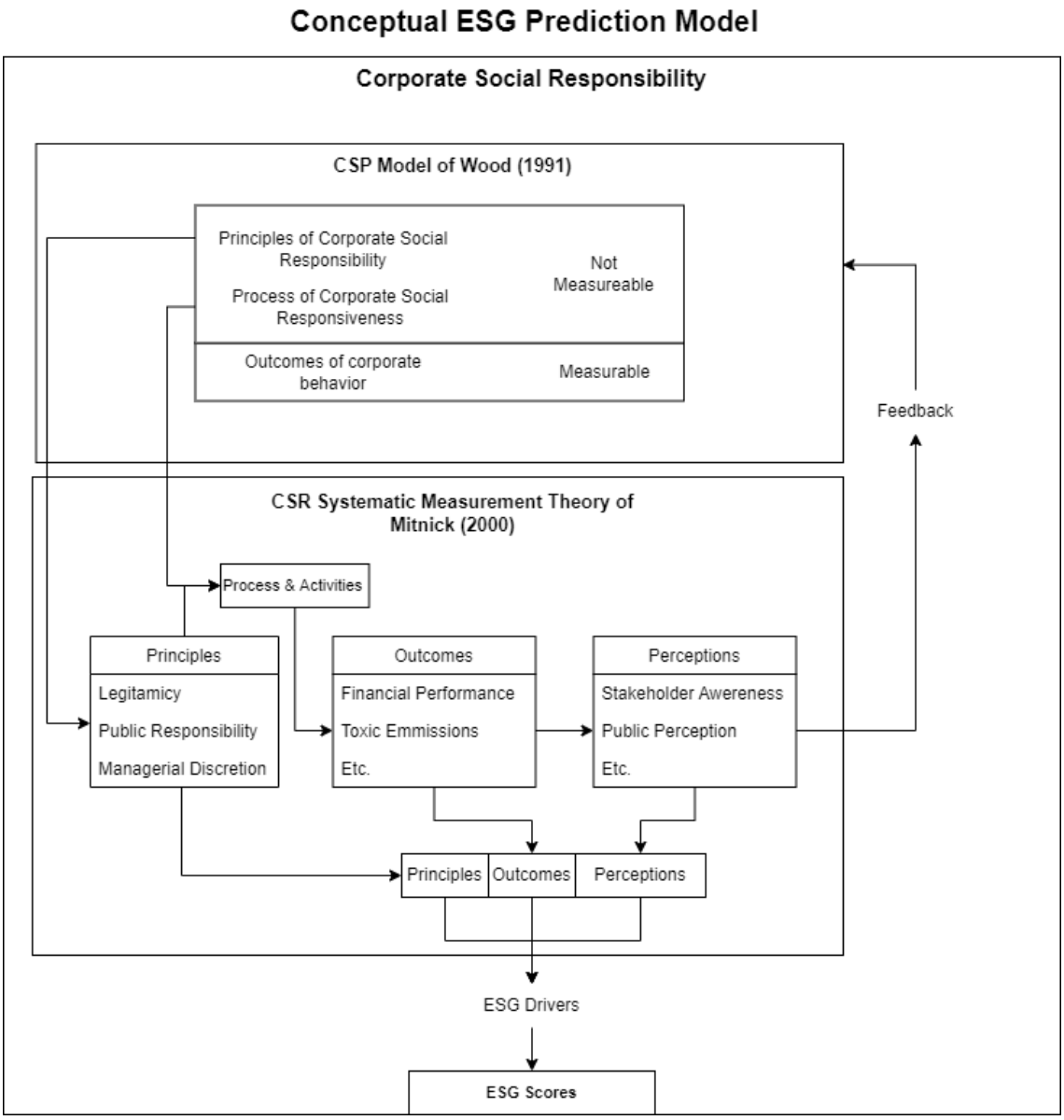


Figure 2: Conceptual ESG Prediction Model

This conceptual model should not be interpreted as a definite model for ESG predictions but rather as a starting point to summarize ESG prediction literature. This is for two reasons. First, ESG predictions is a complex topic, many variables can influence ESG predictions. Second, whilst the literature on CSR and CSP is going back decades, ESG prediction literature does not. Hence, it is expected that this model will and should be updated as more research is completed on the topic.

2.5 Hypothesis and Research Outline

The conceptual model and literature examined in the literature review above help to get a better understanding of the complex topic of CSR to which ESG scores are related. Although the literature has already answered many questions on these topics, many more remain to be researched.

This research aims to contribute to the literature that investigates the relationship between ESG scores and the drivers of ESG scores as to predict ESG scores. In particular this research aims to investigate the relationship between CSP outcomes and the (public) perceptions (reputation) on firms. The following research question will be central to this research:

To what extent do CSP outcomes and public perceptions drive ESG scores of firms?

This research question is based on theoretical literature (Wood (1991), Mitnick (2000) Ali et al. (2017)) that theorizes how to model CSR, CSP and ESG and what drives these factors that influences ESG scores and CSP. This research question is also based on multiple empirical studies (Garcia et al. (2020), Krappel et al. (2021), etc.) that establishes which drivers do and which drivers do not significantly influence CSP and ESG scores.

Based on prior research and theories the following hypothesis are expected to be true:

1. Company size has a significant effect on ESG scores.
Based on all prior research on ESG prediction and on the research of Drempetic et al. (2020) company size seems to be a significant driver for ESG scores. Important to notice here is that it is not clear whether this is a selection bias as a result of the methodology of ESG rating agencies or an actual driver of ESG scores.

2. Public awareness has a significant effect on ESG scores
From the literature on CSR and CSP (Servaes & Tamayo (2013), Mitnick (2000)) it is clear that not only the outcomes but also the perception and awareness of the public (stakeholders) on these outcomes determines the weight certain outcomes have on ESG rating of a company.

3. CFP measures have a significant effect on a firm's ESG rating
The literature research has shown that corporate financial performance metrics and CSR can be positively related, partly because firms that are doing financially well simply have more resources to spend on ESG related topics.

To further structure the research, the following sub-questions will be answered first before the central research question will be addressed:

What is the effect of company size on ESG scores of a firm?

What is the effect of public awareness/perception on the ESG score of a firm?

What CFP measures have an influence on ESG scores?

3 Data & Methodology

This section will provide an overview of where the data used in this research comes from, a description of the most important variables and a short exploratory analysis of the data. Finally, this section will discuss the methodology used to analyse the potential relationships between the variables and to test the hypothesis.

The data for this research is collected from four different data sources. Most important are the ESG scores which are retrieved from the Thomson Reuters database (also known as ASSET4). This data set contains 20 years of annual ESG scores on 9338 companies from 69 countries. Next, the advertising intensity data is retrieved from the Financial Ratios Suite by Wharton Research Data Service (WRDS) this data set contains data on 5217 companies in from which 4119 are from the United Sates. The financial data is retrieved from the Compustat North America database and contains information on 4194 companies from the United States. It should be noted that Compustat also has a Global database containing mostly non-us companies, however, this data has no overlap with the advertising intensity data needed for the research. Hence, all the companies used for this research are based in North-America. Finally, data on the reputation risk and news coverage on companies within the data set is retrieved from the RepRisk, provided via WRDS. This data set contains 11378 companies.

To merge the data properly two different company identifier codes were used, namely: the ISIN (International Securities Identification Number) and the CUSIP code. There are several reasons for this, most importantly some companies are registered with different names in different data sets (i.e. Tesla Incorporated and Tesla Motors) or companies change their name over time (i.e. Facebook to Meta). Company identifiers stay constant over time and are therefore more reliable. Unfortunately, there are different data vendors use different identifiers and as a result, multiple identifiers had to be used in order to properly merge the data, complicating the data merging process.

After merging the data, 1380 companies are left over with data on all variables between 2007 and 2020. In the descriptive statistics and exploratory analysis of this paper this final data set will be further explored before any ESG prediction model is build.

3.0.1 Variable Description

Some variables used in this research have a somewhat more complicated interpretation, therefore a short description of the variables used in this research will be given below.

ESG Data:

The Thomson Reuters (Refinitiv) ESG scores are computed for each company using an extensive data collection and analyzing process. To calculate the score, over 500 data points are collected for each individual company ranging from data on diversity of the workforce to emission which are grouped in 10 categories divided over the three ESG pillars (Environment, Social and Governance). These data points are then translated into percentage rank scoring to get a score between 0 and 100 for each of the three ESG pillars. Next, for the Environment and Social pillar these scores are weighted relative to industry peers to better compare the scores of each company within their industry, since some industries are inherently more "controversial" then other (i.e. weapons manufactures vs food manufactures). Since the governance scores are more dependent on the country where the company is based, it is weighted per country. Finally the weighted scores are combined into a the final ESG score. It should be noted here that Thomson Reuters also offers a so called "combined ESG score" which extends the ESG score from the ASSET4 data with a controversy layer. Although this is related to this research, as controversies can be a good indication for public perception, it is not used in this research because the data was not available to this research.

Some caveats with working with ESG data are already explained in the literature review, however, it should be noted again that although the interpretation of an ESG score is rather simple, the underlying data is complex and subject to multiple calculations, adjustments and interpretations of the analysts. As can be seen from the summary of how the data is collected and processed above. Thus, any analysis and conclusions using ESG data should take this into consideration.

Public/Customer Awareness Data:

Public awareness and perception is not a simple variable one can measure like, for example, market value. Therefore, two variables are incorporated into this research to gauge public awareness and perception.

The first variable is the advertising intensity. Advertising intensity is the advertising expense divided by sales.

$$\frac{Advertising\ Expense}{Sales}$$

The use of advertising intensity is proposed by Servaes & Tamayo (2013) They argue that advertising can help a firm reduce a potential information gap between a firm and its customers. This in turn will lead to a higher customer awareness of a firm's CSR activities and will reward the firm for engaging CSR activity which they base on a number of papers that show that customers take a firm's CSR activities into consideration when making a purchasing decision (for example Barnett (2007)).

As a second variable to gauge public awareness and perception of firm is the reputation risk index provided (RRI) by RepRisk. This is an index that quantifies a firm's reputational risk exposure to ESG related issues. As with the ESG score, the RRI gives a number between 0 and 100, with 100 being the highest reputational risk possible. RepRisk provides three RRI scores, current, peak and trend. For the purpose of this research the current RRI is used which is the current level of media and stakeholder attention of a company related to ESG issues. Although provided on monthly basis, for only the end of the year RRI's are used since the other data sets do not contain monthly data and are recorded at the end of each year (monthly).

As with the ESG the RRI is somewhat more complex in nature. To calculate the RRI RepRisk screens over 100,000 public sources each day in 23 languages to gather data on ESG related risk. The RRI is then calculated based on the reach of the data source, frequency it's mentioned, timing of the ESG related issue, severity and novelty of the ESG related issue. It should be noted here that the RRI only covers ESG related risk, thus it only looks at negative exposure with regard to ESG related issues. Hence, it does not give a full image of the public awareness of ESG issues that a firm might face. After all, the RRI only measures the reputational risk to ESG related incidents, not positive ESG related news.

To clarify, although they might seem similar at first sight, the ESG score and RRI are very different in every dimension with the exception of the scoring scale. The ESG score is computed based on actual performance related to ESG issues (i.e. management pay structure or CO2 emission), the RRI only measures the risk exposure when an ESG related incident occurs and receives attention from the media and stakeholders, it does not measure the validity of the reported allegations.

Firm financial variables:

Most research on ESG score prediction has focused on the relationship between a firm's financial performance and CSR/ESG performance. (I.e. Krappel et al. (2021), Garcia et al. (2020), Schuler & Cording (2006)). Although the results were mixed in some cases, financial performance measures are also included in this research for two reasons. First, including it in the models can extent the earlier works on this topics and provide more conclusive evidence on the relationship between ESG and a the financial performance of a firm. Second, financial performance measures are simple in nature and for a lot of firms publicly available, making analysis easier to interpret than more complex variables such as the RRI. Following the research of Garcia et al. (2020), Ruf et al. (2001) and Waddock & Graves (1997) the following variables will be used as a proxy for a firm's financial performance: the market value of a firm, the return on assets, the return on equity and the debt to equity ratio. A detailed description of the variables used in this research can be found in the appendix.

As described in the in the ESG variable paragraph, industry can explain large variations between companies, its is included as a variable to control for its effect. To classify each industry the Global Industry Classification Standard (GICS) is used of which there are 67 different industries across all the companies in the final data set. Finally, company size is computed in line with the research of Garcia et al. (2020) and is the natural logarithm of the market capitalization of a firm. The reason for using the natural logarithm is because the market capitalization of companies can vary in large degrees with some extreme outliers. Therefore the logarithmic scale is used to scale the data into a more linear pattern, to improve the quality and efficiency of the model.

3.0.2 Descriptive Statistics & Exploratory Analysis

Before the relationship between the variables can be modelled it is important to assess what the data looks like and what statistical models can be used to predict ESG scores. In table 2 an overview of the descriptive statistics van be found. For a more graphic view of the observations, the appendix contains histograms and box plots of all variables used in this research.

It should be noted that ROA, ROE and D/E and advertising intensity all have extreme values that are far away from the mean. A careful approach to extreme values in a sample should always be taken due to the nature of regression analysis, extreme values can disproportionate affect estimators and thus the interpretation of the results. In this data set it seems that D/E and advertising intensity have some extreme values that can both influence the estimator significantly and are far from what

Table 2: Descriptive Statistics

	ESG Score	ROA	ROE	D/E	Advertising Intensity	Size	RRI
Min.	0.50	-4.08663	-142.9534	-3529.106	0.00000	1.5	0.000
1st Qu.	28.01	0.03284	0.0438	0.741	0.00000	840.1	0.000
Median	40.63	0.10520	0.1020	1.557	0.00000	2681.4	0.000
Mean	42.92	0.10253	0.0731	2.958	0.01404	13611.4	9.719
3rd Qu.	56.44	0.16830	0.1758	3.750	0.01260	9011.3	19.000
Max.	95.19	1.47880	24.4338	7781.871	9.54296	1540774.2	74.000
NA's	4851	43	553	25	134	1050	0

Total possible observations: 16472

is expected even in extreme cases. For example a company with an advertising intensity of 9.54 means that a firm in the sample spend 9.54 times the amount of sales (revenue) on advertising and thus spend more time on advertising than there are sales. For the purpose of this research, the data is not adjusted for outliers, because the outliers in the data set can be real scenarios. Moreover, there are only few outliers suggesting that the overall impact of outliers is minimal. However, some robustness checks will be performed to address some outliers or data formats.

It is also clear that, in line with expectation, the distribution of the RRI. The reason this is in line with expectation is because, as mentioned in the variable description section, the RRI only captures ESG related risk incidents at stakeholders and in the media. Hence, if there is no serious ESG related incident, as most firms have, the RRI is 0.

One last important step before building models and analysing their results is to check if there are any correlations between the variables. In 3 the correlation-matrix of all the variables is displayed.

Table 3: Correlation Matrix

	ESG Score	ROA	ROE	D/E	Advertising Intensity	RRI	Size
ESG Score	1.00						
ROA	0.13	1.00					
ROE	0.06	0.21	1.00				
D/E	0.02	-0.02	0.03	1.00			
Advertising Intensity	-0.02	-0.05	-0.02	0.00	1.00		
RRI	0.48	0.08	0.02	0.03	0.02	1.00	
Size	0.58	0.25	0.08	0.03	0.00	0.55	1.00

The correlation-matrix show that there a correlation of .48 between ESG score and RRI and a correlation of .58 between ESG score and Size. All other variables have no substantial correlation with the ESG score. However, there is also a relatively strong correlation between RRI and Size.

Indicating that there might be some information/variation of ESG scores that both variables explain. It is also interesting that there is no substantial correlation between advertising intensity and any other variable, hence it is expected that when running models to predict ESG scores, advertising intensity will not be of substantial explanatory value to ESG scores.

3.1 Methodology

Based on the exploratory analysis a linear regression model is chosen to predict ESG scores based on the variables collected from the various data sources. Moreover, to account for any industry effects fixed-effects estimation will be used instead of a pooled OLS estimation with this panel data. There will be no time-fixed effects so the time dimension of the data will not be used and is thus pooled. As previously mentioned this is done because it is not expected that ESG scores vary significantly over time in the way that, for example, stock market volatility or business cycles do.

In general a fixed-effects model looks as following:

$$Y_{it} = \beta_1 X_{1,it} + ... + \beta_k X_{k,it} + \alpha_i + U_{it}$$

Here α_i is the intercept for each entity state, thus it is the intercept for each industry. β_k is the estimated coefficient and U_{it} is the error term.

To test if the fixed effects model is more efficient than a pooled OLS an F-test will be performed. Where the null hypothesis assumes that the pooled OLS is a better model than the fixed effect model. Hence, if the p-value of the F-test is smaller than the 5% level (.05) the fixed effect model is a better model.

Although previous literature uses more complex models such as machine learning techniques, this research will not. This is to limit the scope of this research and to not further degrade the complicated interpretability of the prediction model, as mentioned in the data section, some variables used in this research already have a rather complex interpretation.

Some caveats of using a linear model estimation technique are of course that it might not capture all the information available in the complex data used for this research, as you assume only linear relationships between the variables. However using log transformations and fixed effects it is expected that the most important information in the data is captured using the models constructed. Moreover, the regression models also enables the testing and interpretation of potential causal

relationships between the variables used in this research that are easier to interpret than some machine learning models.

One estimation technique closely related to the fixed effects estimation methodology is the random effects estimation. The main difference between the random effects and fixed effects estimation is that with random effects variables are allowed to vary to some degree between entities (or time). With fixed effects estimation this is not possible. Allowing for some degree of variation between entities in the case of this study that means between that some degree of variation between industries is allowed, as opposed to no variation across different industries with fixed effect estimation. This can improve the fit and efficiency of the model and is thus preferred to a fixed effects estimation model. A general random effects model looks as following:

$$Y_{it} = \alpha + \beta_1 X_{1,it} + ... + \beta_k X_{k,it} + u_{it} + \epsilon_{it}$$

Here u_{it} is the between-enity error and ϵ_{it} is the within-entity error. To be able to use random effects estimation we need to assume that the individual effect (α_i) in uncorrelated with all the regressors such that $Cov(\alpha_i, x_{it}) = 0$. Thus in the model above u_{it} must be uncorrelated with ϵ_{it}. If this is the case random effects estimation is more efficient than fixed effects estimation. To test whether $Cov(\alpha_i, x_{it}) = 0$ a Hausman test can be performed which assumes this under the null hypothesis. Hence, if the p-value of this test is smaller than the 5% level (.05) the null hypothesis is rejected and a fixed effects estimation is better.

Due to the nature of the random effect estimation technique the coefficients (results) become more difficult to interpret. Therefore, when the Hausman test is performed and a random effect model is deemed more efficient, a fixed effect model will also be performed to compare the results of both test. After all, to make the assumption that the state effects α_i is uncorrelated with any of the explanatory variables is a difficult task.

As a final step in the model selection and building, a Breusch-Pagan test is performed where the null hypothesis states that the variance is homoskedastic. If the null hypothesis is rejected, that is the p-value $< .05$, robust standard errors are required and thus a robust covariance matrix is used to account for the heteroskedasticity.

To test the hypothesis, a specific to general model (SPIC) approach will be used to construct

the model. That means that models will start with the relationship (variable) of interest for each hypothesis and then will be expanded until all variables are in one model. In total nine models will be constructed to test the different hypothesis and to test the overall predictive ability of all the variables used in this research. Below a more detailed explanation of each model will be provided. All models will use industry-fixed effects because, as aforementioned, some industries are inherently more ESG risky, than others. There will be no use of time-fixed effects because it is not expected that ESG scores vary significantly over time for other reasons than those that are already incorporated into the model. Note that they might differ for other cross-sectional reasons not incorporated in any of the models, in which case there would be omitted variable bias (OVB). Moreover, the F-test, Hausmen test and Breusch-Pagan test are performed to test whether the fixed effects estimation model is the optimal model compared to the pooled OLS or random effects model and to test whether robust standard errors are necessary.

3.1.1 Model Specification

The first hypothesis will be tested using equation 1. Here ESG score is regressed using only size as an explanatory variable. As discussed this, first to test whether size has a significant effect on ESG scores and second, if there is a significant effect, to assess the magnitude of this effect. Based on previous literature especially Drempetic et al. (2020), Licari et al. (2021), Garcia et al. (2020).

$$ESG\ Score_{it} = \beta_1 Size_{it} + Industry\ Fixed\ Effects + \epsilon_{it} \tag{1}$$

Next, hypothesis two will be tested using equations 2, 3, 4 and 5. The reason to test all variables separately and then together is to check the individual effects of the variables and their joint effect on the ESG score. As they are all expected indicators of the level of public awareness of a firm. This makes it possible, hypothetically, that there is potentially some multicollinearity at play when using both variables as independent variables. However, the correlation matrix in table 3 indicates that there is unlikely. Finally, the size variable is added because there is some correlation between ESG score and the RRI as can be seen in 3 which can indicate they might partly explain the same variation in ESG scores.

$$ESG\ Score_{it} = \beta_1 Advertising\ Intensity_{it} + Industry\ Fixed\ Effects + \epsilon_{it} \quad (2)$$

$$ESG\ Score_{it} = \beta_1 RRI_{it} + Industry\ Fixed\ Effects + \epsilon_{it} \quad (3)$$

$$ESG\ Score_{it} = \beta_1 Advertising\ Intensity_{it} + \beta_2 RRI_{it} + Industry\ Fixed\ Effects + \epsilon_{it} \quad (4)$$

$$ESG\ Score_{it} = \beta_1 Advertising\ Intensity_{it} + \beta_2 RRI_{it} + \beta_3 Size_{it} + Industry\ Fixed\ Effects + \epsilon_{it} \quad (5)$$

To test the third hypothesis the models 6 and 7 will be constructed. The first is to only test the most common used CFP measures with regards to the ESG score from the literature. Moreover, size is added to test the models used mostly in the ESG CFP prediction literature and because some CFP measures have a correlation with size as can be seen in 3 which can indicate they might partly explain the same variation in ESG scores. This model can be compared to all other "newer" models specified in this research.

$$ESG\ Score_{it} = \beta_1 ROA_{it} + \beta_2 ROE_{it} + \beta_3 D/E_{it} + Industry\ Fixed\ Effects + \epsilon_{it} \quad (6)$$

$$ESG\ Score_{it} = \beta_1 Size_{it} + \beta_2 ROA_{it} + \beta_3 ROE_{it} + \beta_4 D/E_{it} + Industry\ Fixed\ Effects + \epsilon_{it} \quad (7)$$

Finally, as one of the main goals of this paper is to predict ESG scores a final model will be constructed to test whether all variables together are a good model to predict the ESG values. Equation 8 given an indication of what the final model will look like.

$$ESG\ Score_{it} = \beta_1 Size_{it} + \beta_2 ROA_{it} + \beta_3 ROE_{it} + \beta_4 D/E_{it} + \beta_5 RRI_{it} + \beta_6 Advertising\ Intensity_{it} + Industry\ Fixed\ Effects + \epsilon_{it} \quad (8)$$

4 Results

In this section the results of the regressions models constructed will be discussed. First, the results will be discussed per hypothesis. After the results of each individual hypothesis is analyzed the final regression model discussed This model will incorporate all independent variables used to test the other hypothesis to see whether they jointly improve the quality of the predictions.

4.1 The Effect of Size on ESG Scores

As described in the methodology, a fixed-effects model will be constructed to test the potential relationship between a firm's size and its ESG score. However, to test whether this is the optimal model, an F-test and Hausman test are performed to test how whether a pooled OLS or random effect model might be more suitable. As described in the literature section in all research on ESG scores reviewed in this research firm size has the most substantial impact on ESG scores. Moreover, based on the descriptive statistics a Breusch-Pagan Test for heteroskedasticity is performed on the fixed effect model to test whether it is necessary to use robust standard errors. The results of these tests can be found in the appendix in table 8. Based on these test the most appropriate model is the random effect model with no robust standard errors. The results of the regression analysis can be found in table 4. However, as described in the methodology section, the results of the random effect model are somewhat more difficult to interpret so the results of the fixed effect are also displayed in table 4 as the results are easier to interpret. In the appendix, figure 5 displays a visual interpretation of the fixed effects as opposed to a pooled OLS model.

From the p-value of the Breusch-Pagan test it is clear that the alternative hypothesis, namely that the residuals are distributed with equal variance (homoskadastic), is rejected. This means that robust standard errors have to be used to still be able to do inference on the standard error, t-stat and the confidence interval of the estimator. It does however, increase the standard error thus making the model less efficient.

The results of both the random effects and the fixed effects model in table 4 are very similar.

Table 4: Size effect on ESG scores

	Model 1.1	Model 1.2
(Intercept)	-18.46^{***}	
	(1.04)	
Size	7.16^{***}	7.17^{***}
	(0.09)	(0.09)
R^2	0.37	0.35
Adj. R^2	0.37	0.34
Num. obs.	10857	10857
Model	Random Effect	Fixed Effect
Robust Standard Error	No	No

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

The Size parameter only differs by 0.01, moreover, the r-squared and adjusted r-squared only differ .02 and .03 respectively indicating only a small improvement in the fit of the model. The results in table 4 show that size of a firm measured in market capitalization has a significant positive effect on the ESG score of a firm for both model. However, because the size effect is measured as the log base e (natural logarithm) of market capitalization it is somewhat more difficult to interpret. Based on the fixed effect model it means that if the market capitalization increase by ~ 2.718 billion US dollars the ESG score increase by 7.17 points. Based on the robust standard error the model is 95% confident that this relationship is somewhere in the range of [6.64; 7.70]. This is rather large, for example, if taken an extreme case such as Microsoft, which had a market capitalization of $1.540.774.200.000 (1.5 trillion USD) by the end of 2020, the ESG score (solely based on size) would range between 94.6 and 109.7 a range of 15.1. Two limitations of the fixed effect model are displayed here. The first is that a linear model like the OLS fixed effect can predict values that are beyond the possible range of the data. An ESG score of more than 100 is impossible by the way ESG scores are constructed. The second shortcoming is that it is not always very precise, especially when, like in this case, heteroskadasticity needs to be accounted for. A range of 15.1 on a scale of 0 to 100 is rather large. Next, the adjusted r-squared from this model is .34 meaning that approximately 34% of variation in ESG scores in the data is explained by this model. Finally, the results indicate that even with robust standard errors the effect of size is significant on a 1% level.

4.2 The Effect of Public Awareness on ESG Scores

To test the second hypothesis, that public awareness has a significant effect on ESG scores, four models are constructed. One to test the individual effect of advertising intensity, one to test the

individual effect of the reputational risk index, one where there effect together is tested and lastly one where they are both combined and the size variable is added. Again, this is because size has a correlations with the RRI and thus they might (partly) explain the same variation (information) on ESG scores. As with the model for size the same steps are taken to determine the optimal model. That is, pooled, fixed effect or random effect with or without robust standard errors. The results of these test can be found in table 8. The models selected based on these values are displayed in table 5. It is important to note here that just as with the individual size model the random effect method was more appropriate based on the Hausman test. However, to be able to better interpret and compare the results and models a the fixed effect estimation method is also applied to model 3.

Table 5: The Effect of Public Awareness on ESG Scores

	Model 2	Model 3.1	Model 3.2	Model 4	Model 5
Advertising Intensity	−2.45			−2.13	−0.83
	(1.81)			(1.12)	(1.04)
(Intercept)		34.14***			
		(0.90)			
RRI		0.67***	0.67***	0.67***	0.32***
		(0.03)	(0.03)	(0.03)	(0.03)
Size					5.69***
					(0.29)
R^2	0.00	0.25	0.22	0.22	0.38
Adj. R^2	−0.01	0.25	0.22	0.22	0.38
Num. obs.	11543	11621	11621	11543	10787
Model	Fixed Effect	Random Effect	Fixed Effect	Fixed Effect	Fixed Effect
Robust Standard Error	No	Yes	Yes	Yes	Yes

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

The results in table 5 indicate that, as with the size model, the difference between the fixed effect and random effect model with this data set is minimal. The r-squared and adjusted r-squared increase by .03 for both measures of fit when using the random effect as opposed to a fixed effect model. What is more, the coefficient of the RRI stays the same when using either method at .67. Hence interpreting the coefficient results in the same conclusion. The results in table 5 also show that advertising intensity does not explain any variation of ESG scores and even results in a negative adjusted r-squared of -0.01 indicating that it might even do more harm in explaining variation in ESG scores than it will do well. Moreover, the coefficient of advertising intensity is also not significant with a p-value of 0.1747. Hence interpreting the coefficient would not add any

value to this research. The RRI does have a significant coefficient at 0.1% levels that also seems to explain about 22-25% of variation according the r-squared and adjusted r-squared. This means that the RRI does contain some information about ESG scores. The 95% confidence interval of the RRI is [0.61; 0.73]. To put this in perspective General Motors had an RRI of 74 in 2014. Based solely on the RRI the model is 95% certain the ESG score of General Motors in this year would be between 48.81 and 53.29. The actual ESG score was 60.79 in this year, which is quite far of. However, this can be expected with an r-squared of about .22-.25. In fact, due to the coefficient of .67 and the nature of the RRI and ESG data, with a maximum of 100, it is impossible for the RRI to explain all of the variation in ESG scores. A company with a RRI of 100 would only have a ESG score between 63 and 73. As expected by the results of model 2 the combined effect of advertising intensity and the RRI does not improve the quality of the model by any measurement since the RRI and Size have no correlation and advertising intensity has no significant relationship with the ESG score. Except for a loss of some observations the model is equivalent to running just the RRI model. Conversely, by including size as a explanatory variable in model 5 the conclusions and interference based on the previous models does seriously change. The model fit measured in r-squared and adjusted r-squared increases from .22-.25 to .38. Moreover, the coefficient for the RRI more than halves to .32 whilst staying significant at the .1% level. The coefficient of size in this model is 5.69 this is down about 1.48 compared to the model that only uses Size to predict ESG scores. This can suggests that RRI and Size contain some of the same information about ESG scores. However, the standard error, partly due to the need for robust standard errors, increases for size and thus reducing efficiency of the model.

Concluding, advertising intensity in its current form with all the models and methods tested do not have an effect on ESG scores. RRI on the other hand does have a significant effect on ESG scores albeit reduced when size is included in the equation. The results will be further discussed and put into perspective of the literature on CSR and ESG in the discussion section.

4.3 The Effect of CFP Measures on ESG Scores

To test the third and final hypothesis, that CFP measures have an effect on ESG scores, 2 models are constructed. Again, one model with only the CFP measures and one where size is also taken into the equation. To choose to most appropriate method the same step-wise methodology is applied to as before, namely an F-Test to choose determine whether fixed effects are preferred over pooled OLS, thereafter a Hausman Test to compare the fixed and random effects models and finally the

Breusch-Pagan test is performed to determine whether robust standard errors are necessary. The results for model 6 and 7 can be found in table 8. As expected, the fixed effects model is always preferred to the pooled OLS. However, the Hausmen test results in a p-value of 0.01837, which means the p-value is only significant at the 5% level and not at the 1% or .1% level. Therefore, as with earlier models, both the fixed effect and random effect model are displayed, to compare the results. The results of model 6 and 7 can be found in table 6.

Table 6: The Effect of CFP Measures on ESG Scores

	Model 6.1	Model 6.2	Model 7
(Intercept)	41.01***		
	(1.19)		
ROA	15.78*	15.66*	−8.69**
	(6.78)	(6.84)	(3.18)
ROE	0.32	0.31	0.13
	(0.19)	(0.19)	(0.09)
D/E	0.01	0.01	0.00
	(0.00)	(0.00)	(0.00)
Size			7.42***
			(0.27)
R^2	0.05	0.01	0.35
Adj. R^2	0.05	0.01	0.35
Num. obs.	11236	11236	10494
Model	Random Effect	Fixed Effect	Fixed Effect
Robust Standard Error	Yes	Yes	Yes

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

The results in table 6 show that, as with the models discussed before, the results of the fixed and random effect models are very similar. Again, the biggest difference is in the r-squared. The r-squared and adjusted r-squared of the random effect model is 5 times higher than the ones from the fixed effects model at .05. However, both models explain very little of the variation in ESG scores with the r-squared and adjusted r-squared of .05 and .01 for the random effect and fixed effect models respectfully. Moreover, the return on assets (ROA) is the only explanatory variable of all the CFP measures that significant at least the 5% level. Moreover, the standard error of the ROA is relatively large compared to the coefficient, this also results in a 95% confidence interval of [2.24; 29.07] for the fixed effect model, a range of 27.46, which is rather large. Based on the data from the descriptive statistics this means that when applying model 3 the ROA based ESG score can range between [-118.78; -9.15] at minimum and [3.31; 42.99] at maximum. Next, the results in table 6 also show that return on assets and the debt to equity ratio are not significant at the

.1, 1 and 5% level. And even if they were significant, their coefficient are small and with the data used in this research it would not be fully able to explain the variation is ESG scores. Surprisingly when the variable size is taken into the equation in model 7 the coefficient of ROA changes from a positive to a negative sign and becomes -8.69 whilst becoming more significant at even the 1% level as opposed to the 5% level. Indicating that an increase in the return on assets would decrease the ESG score of a firm. What is more, the size coefficient of 7.42 is not far off from the one in model 1 (7.17). However, this can also be because the explains very little variation of ESG scores that is not already explained by size, as the r-squared and adjusted r-squared show an increase compared to the model 6.1 and 6.7 from .01-.05 to .35 it is about the same as with model 1 (.34-.37).

To conclude, the CFP measures used in this research have little to none explanatory power over the variation in ESG scores. The size variable together with the RRI variable seems to be the only variables that are both significant and add fit the model measured in r-squared.

4.4 A Combined Model for ESG Score Prediction

Finally, as one of the main goals of this research is to better predict ESG scores, two models are constructed. One with all the variables used in the previous models and one with all variables used in this previous models except for advertising intensity. This is because advertising intensity can even negatively impact the fit of the model and therefore it is interesting to see what happens to the overall performance of the model when it is not included. As with the previous models, to decide between a pooled OLS, fixed or rand effect model the F-test and Hausman test are performed. Finally to check whether robust standard errors are required the Breusch-Pagan is performed. The results of these tests can be found in table 8. Based on these test, fixed effects methodology is deemed most appropriate for models 8 and 9. However, the Breusch-Pagan test p-value of model 9 is 0.02896 and thus only significant at the 5% level. Although this is sufficient, the results of the random effect model are also added to table 7.

The results in table 7 are very similar to the results in table 5, especially model 5. Based on the results of the other models in this research this does not come as a surprise. After all, all the CFP measures used in this research had no significant nor substantial effect on ESG scores. The only significant coefficients in both model 8 and 9 are the RRI and the Size coefficients. The RRI coefficient of .31 only differs about .01 compared to model 5 and the standard error only increases by .02 compared to model 5. As both models use the same methodology, namely fixed effects and robust standard errors the interpretability stays the same. Moreover, the size variable also deviates

Table 7: Combined Model for ESG Prediction

	Model 8	Model 9.1	Model 9.2
(Intercept)			-10.69^{***}
			(2.11)
Advertising Intensity	-1.14		
	(1.05)		
RRI	0.31^{***}	0.31^{***}	0.31^{***}
	(0.03)	(0.03)	(0.03)
ROA	-6.43^{*}	-5.06	-4.79
	(2.90)	(2.95)	(2.93)
ROE	0.15^{*}	0.14	0.14
	(0.07)	(0.08)	(0.08)
D/E	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Size	5.92^{***}	5.90^{***}	5.87^{***}
	(0.30)	(0.30)	(0.30)
R^2	0.38	0.38	0.40
Adj. R^2	0.38	0.38	0.40
Num. obs.	10449	10494	10494
Model	Fixed Effect	Fixed Effect	Random Effect
Robust Standard Error	Yes	Yes	Yes

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

only slightly compared to model 5. In model 5 the size coefficient is 5.69, in model 8, 9.1 and 9.2 it is 5.92, 5.90 and 5.87 respectfully. Interestingly the random effect model has a higher r-squared and adjusted r-squared suggesting the model is a better fit for the data. However, the improvement is only minimal at .02 point compared to model 9.1. Next, the ROA and ROE in model 8 are significant at the 5% level, however, when the advertising intensity is removed from the equation their p-values drops to 0.08578 for ROA and 0.08507 for ROE in model 9.1 and 0.10138 for ROA and 0.06759 for ROE for model 9.1. Finally, both ROE and D/E have a small coefficient indicating that even if they were significant, the impact of a change in either would not result in large changes in the ESG score of a firm.

To conclude, the final, combined, models of this research bear no surprises based on the results of the earlier models. They are very comparable to model 5, the model where only the RRI and Size were included. Hence, based on the results combining all variables into one model does not improve estimation or predictability of ESG scores.

4.5 Robustness Tests

Before the results above will be discussed and a conclusion can be drawn a two robustness checks are performed. These robustness checks will focus on the CFP measures. The first robustness check performed is on the debt to equity ratio. A concern one might have is that the data includes negative debt to equity ratios. This indicates that a company has negative equity and thus most likely will go bankrupt. To test whether if this significantly change the data both models 6 7 and 8 are rerun to see if there are notable differences. Table 9 in the appendix contains the results of these regression, note that the same methodology with regards to model selection was applied as with all the other models. The results show that, in model 6, the debt to equity ratio becomes significant. However, the r-squared goes go down from .05 to .02 and when size is accounted for, the debt to equity ratio becomes insignificant again. Moreover, the magnitude of the coefficients stays small and thus, unless you are dealing with extreme outliers, the effect of the debt to equity ratio stays minimal.

The second robustness check is on the relationship between ROA, ROE and ESG scores. Since the correlation table (see table 3) indicates there is some correlation between the ROA and the ROE. This indicates that they might partly explain the same variation in ESG scores. Therefore a robustness test is performed by rerunning model 6, 7 and 8 with either the ROA or the ROE left out to see if only using one of the two alters the conclusions from the main results. The results of these 6 models can be found in the appendix in table 10. Again the same methodology with regards to the model selection is applied as with all the previous models. The results in table 10 indicate that the ROE does not become significant, moreover, the results also show that the ROA becomes insignificant in model 8 when ROE is not included in the equation. However, the r-squared in both model 8 without the ROE and model 8 with both the ROA and the ROE stays the equal. Hence, the conclusions drawn from the main results do not alter when using only the ROA or the ROE instead of both.

To conclude, although there are some there are some changes to the models and interpretation of the results when performing these robustness tests, they are either of small magnitude, insignificant or do not change fit of the model.

5 Discussion & Conclusion

In this section the results of this research will be discussed and put into perspective of what was expected based on the literature and exploratory analysis. After the results have been discussed, the central research question will be answered in the conclusion.

The first hypothesis tested in this research is:

1. Company size has a significant effect on ESG scores.

This hypothesis is based on the empirical literature on ESG prediction and CSP outcomes. Especially the work of Garcia et al. (2020) and Licari et al. (2021). Both found a significant effect of the size on the ESG score. The results of this paper further strengthen the evidence for this relationship as size is found to be a significant estimator for ESG score. Both in terms of statistical significance and magnitude. Moreover, compared to the linear regression performed in the work of Garcia et al. (2020) both the r-squared and adjusted r-squared are improved in this research. .37 compared to about .27 in the research of Garcia et al. (2020). One possible reason for difference is the use of industries. In this research 67 different industries are included, whereas Garcia et al. (2020) only use 6 industry classifications in their research. Moreover, the fixed-effect estimation also outperformed the linear regression of Licari et al. (2021) in terms of r-squared (.37 in this research compared to .31 in theirs). Although the coefficient in this research does deviate somewhat from the one in Garcia et al. (2020), this is probably because not all variables are accounted for, leading to an upward bias, 7.16 in this research compared to 5.66 in their research. However, when accounting for financial performance measures this difference in coefficients becomes remarkably small.

The second hypothesis tested in this research was:

Public awareness has a significant effect on ESG scores

The effect of public awareness was measured by two variables, advertising intensity and the reputational risk index (RRI). Here, advertising intensity was used as a measurement to decrease any information gap between a firm and its consumers, a theory proposed in the work of Servaes & Tamayo (2013). Their results suggest that without customer awareness, measured as advertising intensity, firms cannot be rewarded for CSR related activities. Based on their results the hypothesis was drawn that customer awareness could also contain information on and or public perception influence the public perception on a company. Moreover, based on the theory of Mitnick (2000),

who argued that for it not only matters how well a firm achieves on corporate social performance, it also matters to what extent the public believes these achievements. Based on the theory of Servaes & Tamayo (2013) and Mitnick (2000) it was hypothesised that a higher advertising intensity could decrease any potential information gap between the firms CSP and the public perception of these a firm's CSP and hence, effect the ESG score as public perception can be seen as a feedback for the principles of CSR (see the conceptual ESG framework in figure 2). The results of this research indicate that the advertising intensity does not contain information on the variation of ESG scores. Comparing the result of this study with that of Servaes & Tamayo (2013) is somewhat tricky since Servaes & Tamayo (2013) did not regress advertising intensity on ESG scores but combined it with CSR measurements to find whether companies with higher public awareness are more rewarded for their corporate social performance for which they did find evidence. There are multiple explanations for why advertising intensity has no significant effect on ESG scores, most prominent, it might simply not contain any information on the variation of ESG scores. Another possible reason is that it takes time for public awareness and perception to have an effect on ESG score. After all, it can take time before the public perception and awareness is translated to a change in ESG score as it first has to follow the feedback loop, described in the conceptual framework, from public perception back to CSP outcomes and subsequent ESG score Hence using the same year observation for both advertising intensity and ESG score might not be optimal. Lastly, firms might simply not use advertising to communicate there corporate social performance outcomes.

The next variable used to test hypothesis two was the reputational risk index. This variable measures the reputational risk exposure specifically with regards to ESG related topics. The theoretical base for adding the reputational risk index is the same as theory used for the advertising intensity. However, they do measure the public awareness in different ways. Where the advertising intensity measures the public awareness through the theory that advertising could decrease any information "gap" or asymmetry between the firm and stakeholders, the RRI measures it in a more direct method by actually scanning the through the internet and media for articles and documents that could indicate reputational risk related to ESG related topics. Since this variable is, as this research could find, not used in ESG score/rating related research it is somewhat harder to compare the results of this research with other literature. None the less, the results do suggest that there is some form information on the variation in ESG scores in the RRI and thus suggest that public awareness might influence the ESG score. Although the it is impossible to explain all variation of the ESG score with the RRI, the impact of the RRI can be substantial when a company has a

high RRI. In the most extreme case, a RRI of 100 can lead to an increase of 31 point of the ESG score. This suggest that more awareness, even if the perception is negative, increases the ESG score.

The third and final hypothesis tested in this research was:

CFP measures have a significant effect on a firm's ESG rating

This hypothesis was build on a larger literature base, both theoretical and practical. In particular the work of Garcia et al. (2020), Waddock & Graves (1997) and Wang et al. (2016) who all find mostly positive significant relationships between a firm's financial performance measurements and corporate social performance. The results in this research give a somewhat more mixed view on the relationship between a firm's financial performance measurements and its corporate social performance. For starters, this research found no significant relationship between the debt to equity ratio and the ESG score of a firm. Garcia et al. (2020) also found no significant relationship between the debt to equity ratio and the ESG score. However, although the sign for the ROA is the same in both this research and the one of Garcia et al. (2020), the magnitude of the coefficient in the final model (8) differs strongly (-46 in their research versus -6.43 in this research). In none of the models, both in the main results as in the robustness tests, does estimated coefficient of the ROA ever reaches a level close to -46, nor does it reaches the same level of significance. The reason for this difference might be found in the data en methodology used. For example. differences in the industries used as control variables. In this research 64 different industries are used to as control variables, in the research of Garcia et al. (2020) only 5 specific industries are used and all other industries are controlled for with only one variable. Next, there are significant differences in the results of this research compared to those of Waddock & Graves (1997). Most prominently is the relationship between the CSP and the ROA. Waddock & Graves (1997) found a significant positive relationship between the CSP and the ROA of a firm. That is the opposite of what the results in this research suggest. This apparent difference can be due to a number of reasons. For starters, the CSP of a firm in this research is measured by the ESG score of a firm. In the research of Waddock & Graves (1997) it is measured with a CSP index they constructed using data from the KLD database. Moreover, Their data set included data from before 2000. Whilst in this research only data from after 2000 is used, hence there is no data overlap. Finally, Waddock & Graves (1997) do not seem to control for industry, something all models in this research do control for.

5.1 Conclusion

Environment, social and governance (ESG) scores are increasingly important in investment decisions making and are an attempt to measure the corporate social performance of firms. However, due to the inherent nature of the ESG and CSR topics these scores are time consuming and subsequently costly to make. This paper tried to build upon the early attempts in the literature to predict ESG scores to, at the very least, reduce these costs and speed up early indications of what the ESG score of an unrated firm might be. The main goal of this research to further extent the literature on ESG prediction. Since ESG and CSR are complex and dynamic topics this research tried to focus on the relationship between public perception and awareness and ESG scores. In this research, the central research question was:

To what extent do CSP outcomes and public perceptions drive ESG scores of firms?

Based on the literature review and conceptual model three hypothesis and sub question were formulated to help answer the central research question. The first sub question was on the relationship between size and ESG scores. In all models ran in this research size had a significant effect on the ESG score, both in terms of statistical significance and magnitude. Moreover, the results were also similar to that of earlier research.

The second sub question addressed the relationship between ESG scores and public awareness. The results from this research suggest that there might be some relationship, however, the significance of the relationship is not very strong. Both in terms of statistical significance and magnitude. The RRI is significant, and the coefficient indicates that it can explain a substantial part of the variation of the ESG scores. Advertising intensity does not explain any variation in ESG scores, in none of the main results.

The third and final sub question addressed the relationship between ESG scores and a firm's financial performance. The results from this research show a mixed results. Only the ROA has some explanatory power over the ESG score variation, however, when advertising intensity is removed from the model the ROA becomes insignificant. Moreover, the sign for the estimated coefficient of the ROA is negative, indicating a negative relationship, the opposite of what earlier research on the relationship between a firm's FP and CSP found.

These results suggest that public awareness measured in the RRI does have an effect on the ESG scores of firms. Moreover, it can explain a substantial part of a firm's ESG score. However,

the low r-squared of all the models indicate that there is still a large part of variation not explained. Hence including more explanatory variables might alter the explanatory power of the RRI. Next, public/consumer awareness measured with advertising intensity does not have a significant effect on ESG score. The results of this research also provided additional evidence for the hypothesis that size has a significant effect on ESG scores. Finally, this research provided mixed results for the relationship between the financial performance measurements of firms and their ESG score, the ROA can in certain models have a significant both statistically and in terms of magnitude, however, this does not hold in all models ran, making it difficult to draw a conclusion on its effect on ESG scores.

5.1.1 Implications

The results of this research have several implications for companies, academics and society. First, the results imply that having a higher reputational risk exposure on ESG related topics does significantly impact the ESG score. Moreover, a higher reputational risk exposure improves the ESG score. Suggesting that any attention related to a firm's ESG related performance is better than no attention when it comes to ESG scores. Next, the results also imply that spending more on advertising does not improve the ESG score. It should be noted here however that this research did not made a distinction between ESG related advertising and non-ESG related advertising. The results of this study also imply, as in in previous literature, that size is an important factor when it comes to ESG scores. Ceteris paribus, the higher the market capitalization of a firm the higher its ESG score. Implying that larger companies do better in terms of corporate social responsibility than smaller companies. Why this is the case is a topic for further research. Finally, the results of this research imply that not all financial performance measurements significantly impact the ESG score of a firm, or in other words, contain any information about the ESG score variation of firms. This implies that the financial performance of a firm is not necessarily important when it comes to its CSP and subsequent ESG score.

These results also proved to be robust to several robustness tests regarding changes in the CFP measures.

6 Limitations & Suggestions for Further Research

Though this research tried to be as extensive as possible, there were some limitations. The most prominent limitations will be discussed bellowed and followed by suggestions for further research into ESG score prediction.

One of the most prominent limitations of this research is the data. Data forms a limitation in two ways for this research. First, although one of the goals of ESG prediction is to overcome some of data shortage problems, a good qualitative data set is required to estimate and build models. Unfortunately when all variables from the different data sets and data vendors were merged, only data from North American companies was left. Thus the conclusions of this research are only be valid for North American companies. The second limitation of the data used in this research is the complexity of some of the variables used. Both the reputational risk index and the environment, social and governance score are variables that try to capture complex and sometimes dynamic processes. Interpretation of the results of these variables becomes therefore somewhat more difficult. A second limitation of this research is the methodology used. As already mentioned many times in this research, some variables have more complex and dynamic processes on which they are based. Therefore, it might be more suitable to use more complex or advanced models than the fixed effects or random effects estimation models. One important reason for using this was to improve the interpretability of the results and to not further complicate the research than was necessary. However, the results of this research do show a lower data fit than literature that used more advanced methods for predicting ESG scores.

It should come as no surprise that the first suggestion for further research is to use data from different countries to assess how the conclusions of the results hold in other countries, improving the usability of a ESG prediction model. Next, future research could look into other methodology. Especially the methodology of rough sets used in Garcia et al. (2020) is strongly recommended. Their argument that ESG score prediction does not have to be point accurate to be useful. Rather, a model that can predict whether the ESG score falls within a certain range (for example, high, medium low) might be more useful and a better and more realistic goal than a model that perfectly predicts exact ESG score values. Another suggestion for further research is to take the time dimension into account when predicting ESG scores. More specifically take the time dimension into consideration when using variables that try to capture the public awareness and/or perception as

there might be a lag in the time it takes before a change in public perception and/or awareness has an effect on the ESG score. Finally, future research could search and use for other variables that might contain information about the public perception and awareness.

7 Bibliography

References

Ackerman, R. W. (2013). *The social challenge to business*. Harvard University Press. Retrieved from https://doi.org/10.4159/harvard.9780674187771 doi: doi:10.4159/harvard.9780674187771

Ali, W., Frynas, J. G., & Mahmood, Z. (2017). Determinants of corporate social responsibility (csr) disclosure in developed and developing countries: A literature review. *Corporate Social Responsibility and Environmental Management*, *24*(4), 273–294.

Barnett, M. L. (2007). Stakeholder influence capacity and the variability of financial returns to corporate social responsibility. *Academy of management review*, *32*(3), 794–816.

Barnett, M. L., & Salomon, R. M. (2006). Beyond dichotomy: The curvilinear relationship between social responsibility and financial performance. *Strategic management journal*, *27*(11), 1101–1122.

Boffo, R., & Patalano, R. (2020). Esg investing: Practices, progress and challenges. *Éditions OCDE, Paris*.

Bowen, H. R. (1953). *Social responsibilities of the businessman*. University of Iowa Press.

Business Roundtable. (2019). *Business roundtable redefines the purpose of a corporation to promote 'an economy that serves all americans'.* https://www.businessroundtable.org/business-roundtable-redefines-the-purpose-of-a-corporation-to-promote-an-economy-that-serves-all-americans. (Accessed: 19-11-2021)

Carroll, A. B. (1999). Corporate social responsibility: Evolution of a definitional construct. *Business & society*, *38*(3), 268–295.

Carroll, A. B. (2008). A history of corporate social responsibility: Concepts and practices. *The Oxford handbook of corporate social responsibility*, *1*.

Chaffee, E. C. (2017). The origins of corporate social responsibility. *U. Cin. L. Rev.*, *85*, 353.

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? implications for managers, investors and strategy researchers. *Strategic Management Journal*, *37*(8), 1597–1614.

Chatterji, A. K., Levine, D. I., & Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility? *Journal of Economics & Management Strategy*, *18*(1), 125–169.

Corroll, A. (1979). A three-dimensional conceptual model of corporate social performance. *Acad. Manag. Rev*, *4*, 497–505.

Crisóstomo, V. L., de Souza Freire, F., & De Vasconcellos, F. C. (2011). Corporate social responsibility, firm value and financial performance in brazil. *Social responsibility journal*.

Damodaran, A. (2019). *From shareholder wealth to stakeholder interests: Ceo capitulation or empty doublespeak?* `https://aswathdamodaran.blogspot.com/2019/08/from-shareholder-wealth-to-stakeholder.html`. (Accessed: 19-11-2021)

Davis, K. (1973). The case for and against business assumption of social responsibilities. *Academy of Management journal*, *16*(2), 312–322.

Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the esg score: Corporate sustainability ratings under review. *Journal of Business Ethics*, *167*(2), 333–360.

D'Amato, V., D'Ecclesia, R., & Levantesi, S. (2021). Esg score prediction through random forest algorithm. *Computational Management Science*, 1–27.

Freeman, R. E. (2010). *Strategic management: A stakeholder approach*. Cambridge university press.

Friedman, M. (2007). The social responsibility of business is to increase its profits. In *Corporate ethics and corporate governance* (pp. 173–178). Springer.

Garcia, F., Gonzalez-Bueno, J., Guijarro, F., & Oliver, J. (2020). Forecasting the environmental, social, and governance rating of firms by using corporate financial performance variables: A rough set approach. *Sustainability*, *12*(8), 3324.

Gibson Brandon, R., Krueger, P., & Schmidt, P. S. (2021). Esg rating disagreement and stock returns. *Financial Analysts Journal*, *77*(4), 104–127.

The global risks report 2021. (2021). World Economic Forum.

Gruber, J. (2013). *Public finance and public policy* (5th ed.). Macmillan.

Hart, T. A., & Sharfman, M. (2015). Assessing the concurrent validity of the revised kinder, lydenberg, and domini corporate social performance indicators. *Business & Society*, *54*(5), 575–598.

Harvey S. Rosen, T. G. (2014). *Public finance: Global edition* (10th ed.). McGraw-Hill Eduction.

Hu, Y., Chen, S., Shao, Y., & Gao, S. (2018). Csr and firm value: Evidence from china. *Sustainability*, *10*(12), 4597.

Jo, H., & Harjoto, M. A. (2011). Corporate governance and firm value: The impact of corporate social responsibility. *Journal of business ethics*, *103*(3), 351–383.

Kotsantonis, S., & Serafeim, G. (2019). Four things no one will tell you about esg data. *Journal of Applied Corporate Finance*, *31*(2), 50–58.

Krappel, T., Bogun, A., & Borth, D. (2021). Heterogeneous ensemble for esg ratings prediction. *arXiv preprint arXiv:2109.10085*.

Licari, J., Loiseau-Aslanidi, O., Piscaglia, S., & Gonzalez, B. S. (2021). *Esg score predictor: Applying a quantitative approach for expanding company coverage.* `https://www.moodysanalytics.com/-/media/article/2021/esg-score-predictor.pdfl`. Moody's. (Accessed: 25-11-2021)

Makni, R., Francoeur, C., & Bellavance, F. (2009). Causality between corporate social performance and financial performance: Evidence from canadian firms. *Journal of Business Ethics*, *89*(3), 409–422.

Mitnick, B. M. (2000). Commitment, revelation, and the testaments of belief: The metrics of measurement of corporate social performance. *Business & Society*, *39*(4), 419–465.

Mitnick, B. M. (2003). Credible testaments, property, and the role of government. In *The fundamental interrelationships between government and property* (pp. 179–190). Routledge.

Post, J., & Preston, L. E. (2012). *Private management and public policy: The principle of public responsibility*. Stanford University Press.

Ruf, B. M., Muralidhar, K., Brown, R. M., Janney, J. J., & Paul, K. (2001). An empirical investigation of the relationship between change in corporate social performance and financial performance: A stakeholder theory perspective. *Journal of business ethics*, *32*(2), 143–156.

Schuler, D. A., & Cording, M. (2006). A corporate social performance–corporate financial performance behavioral model for consumers. *Academy of management Review*, *31*(3), 540–558.

Semenova, N., & Hassel, L. G. (2015). On the validity of environmental performance metrics. *Journal of Business Ethics*, *132*(2), 249–258.

Sen, S., & Bhattacharya, C. B. (2001). Does doing good always lead to doing better? consumer reactions to corporate social responsibility. *Journal of marketing Research*, *38*(2), 225–243.

Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management science*, *59*(5), 1045–1061.

Sheehy, B. (2015). Defining csr: Problems and solutions. *Journal of business ethics*, *131*(3), 625–648.

Soana, M.-G. (2011). The relationship between corporate social performance and corporate financial performance in the banking sector. *Journal of business ethics*, *104*(1), 133–148.

Van de Velde, E., Vermeir, W., & Corten, F. (2005). Corporate social responsibility and financial performance. *Corporate Governance: The international journal of business in society*.

Waddock, S. A., & Graves, S. B. (1997). The corporate social performance–financial performance link. *Strategic management journal*, *18*(4), 303–319.

Wang, Q., Dou, J., & Jia, S. (2016). A meta-analytic review of corporate social responsibility and corporate financial performance: The moderating effect of contextual factors. *Business & Society*, *55*(8), 1083–1121.

Wartick, S. L., & Cochran, P. L. (1985). The evolution of the corporate social performance model. *Academy of management review*, *10*(4), 758–769.

Weber, J., & Wasieleski, D. M. (2018). *Corporate social responsibility*. Emerald Group Publishing.

Wood, D. J. (1990). *Business and society*. Scott, Foresman/Little, Brown Higher Education Glenview, IL.

Wood, D. J. (1991). Corporate social performance revisited. *Academy of management review*, *16*(4), 691–718.

Wood, D. J. (2010). Measuring corporate social performance: A review. *International journal of management reviews*, *12*(1), 50–84.

8 Appendix

8.1 Abbreviations

CSR: Corporate Social Responsibility

CSP: Corporate Social Performance

CFP: Corporate Financial Performance

RRI: Reputation Risk Index

FP: Financial Performance

IM: Issues Management

ESG: Environment, Social & Governance

GLM: Generalized Linear Model

8.2 Variable Description

$$\text{SIZE} = \log\left(\text{Market Capitalization}\right)$$

$$\text{ROA} = \frac{\text{Net Income}}{\text{Total Assets}}$$

$$\text{ROE} = \frac{\text{Net Income}}{\text{Shareholders Equity}}$$

$$\text{D/E} = \frac{\text{Total Debt}}{\text{Shareholders Equity}}$$

8.3 Descriptive Statistics Expanded

Figure 3: Histogram Of Variables Used

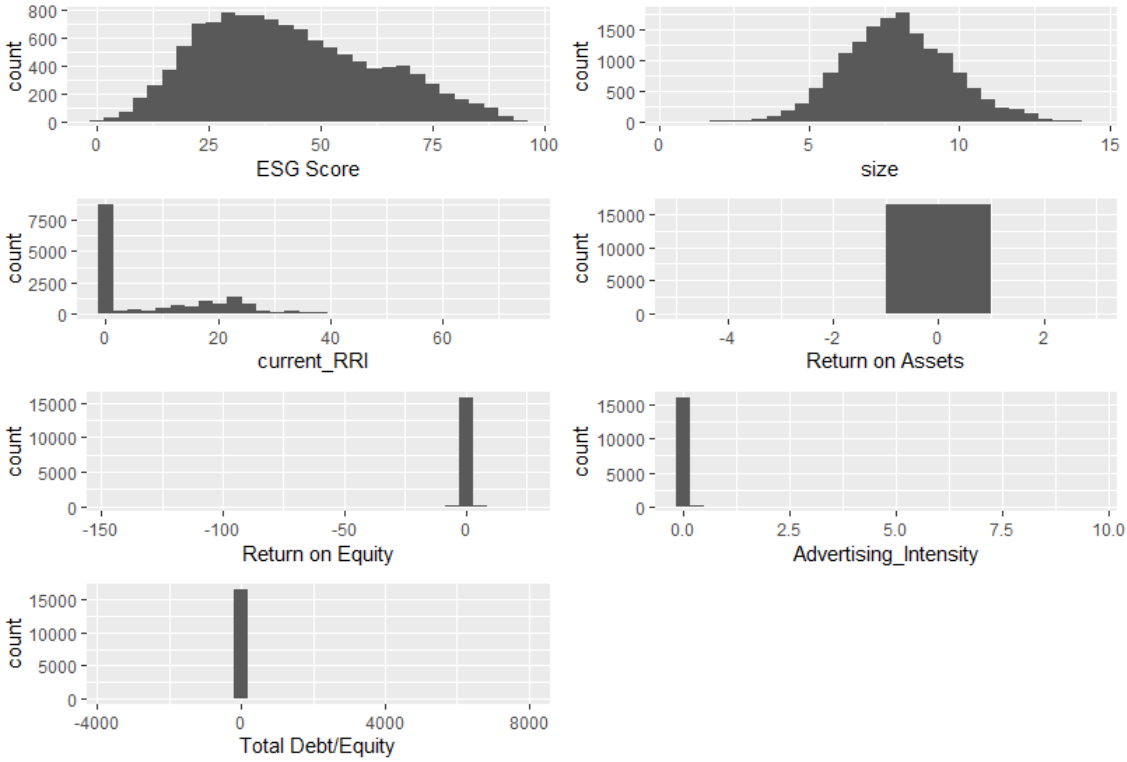
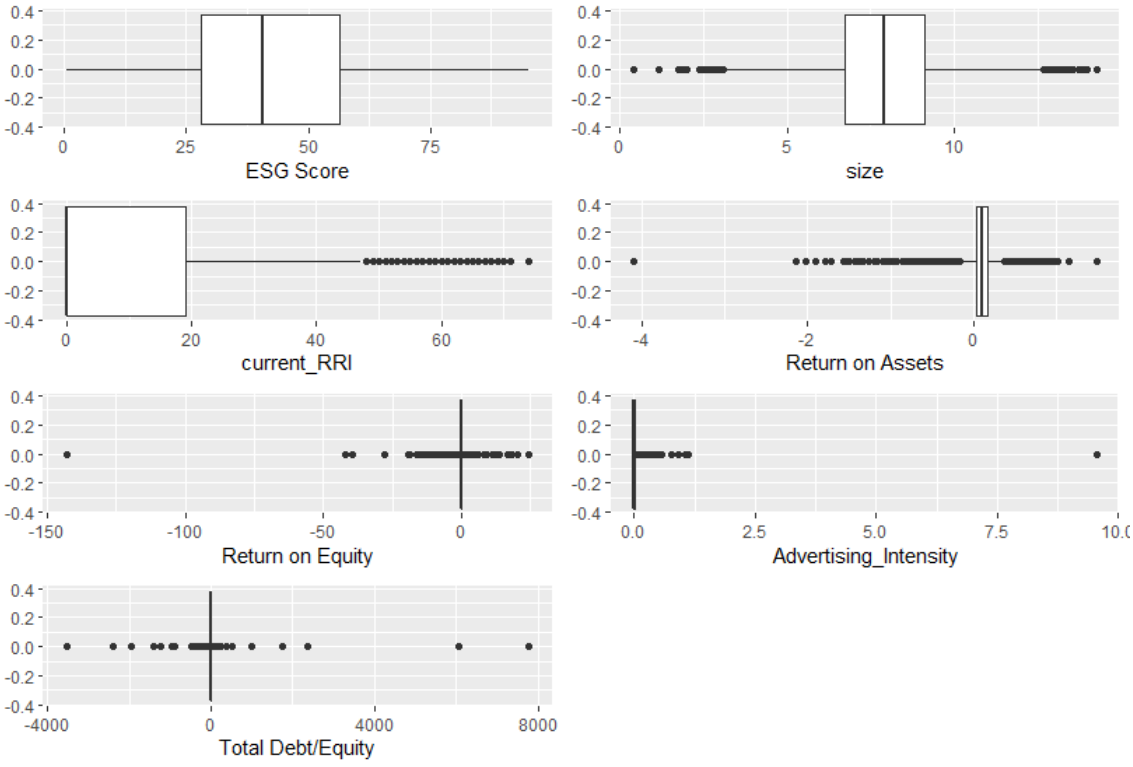
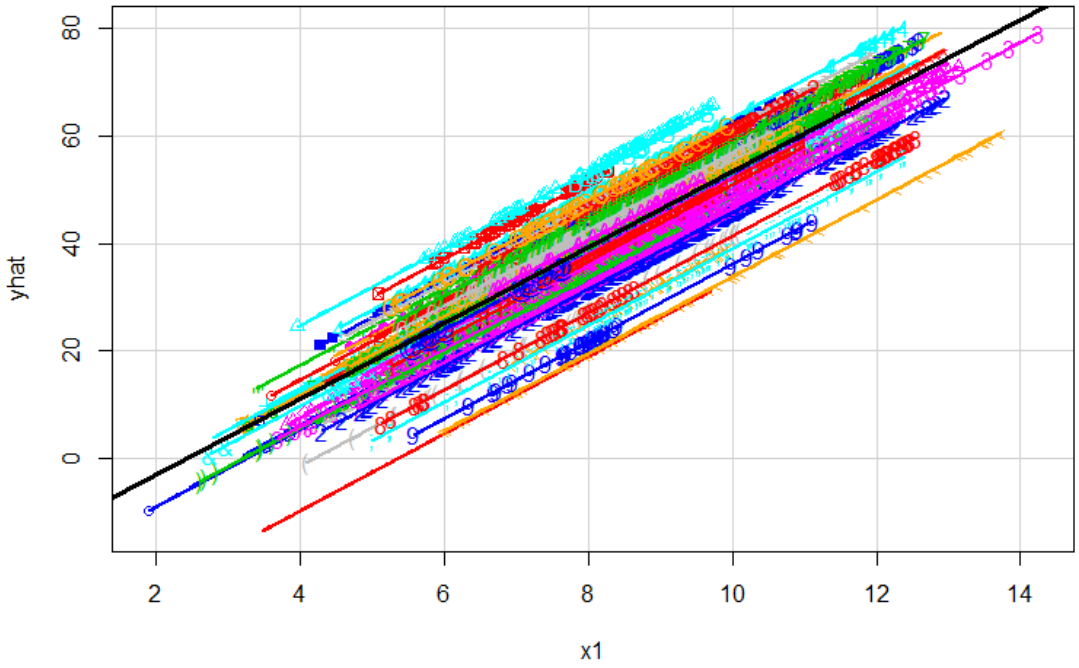


Figure 4: Box Plots Of Variables Used



8.4 Regression Output

Figure 5: Visual Representation of Fixed Effects: Size



This figure illustrates the industry fixed effects on size. Each colored line is a linear regression for each individual industry, the black line is a pooled OLS regression.

8.5 Robustness Tests

Table 8: Model Selection and Robustness Tests

	Breusch-Pagan Test (P-Value)	**F-Test (P-Value)**	**Hausman Test (P-Value)**
Model 1	9.615492e-18	2.2e-16	**0.1131**
Model 2	0.2735	2.2e-16	0.002517
Model 3	2.305038e-30	2.2e-16	0.2156
Model 4	6.611866e-31	2.2e-16	2.2e-16
Model 5	4.814512e-18	2.2e-16	1.246e-09
Model 6	2.2e-16	2.2e-16	**0.6057**
Model 7	1.987e-13	2.2e-16	0.01837
Model 8	9.554e-15	2.2e-16	2.339e-05
Model 9	2.476891e-13	2.2e-16	0.02896

Table 9: Robustness Debt to Equity Ratio

	Model 6	Model 7	Model 8
(Intercept)		-19.64^{***}	
		(2.11)	
ROA	13.00	-8.19^{*}	-6.62^{*}
	(7.59)	(3.41)	(3.15)
ROE	2.16^{*}	0.06	0.29
	(1.03)	(0.44)	(0.45)
D/E	0.07^{*}	0.04	0.02
	(0.04)	(0.03)	(0.02)
Size		7.41^{***}	5.94^{***}
		(0.27)	(0.31)
Advertising Intensity			-1.17
			(1.06)
RRI			0.31^{***}
			(0.03)
R^2	0.02	0.37	0.38
Adj. R^2	0.01	0.37	0.38
Num. obs.	11150	10417	10372

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 10: Robustness ROA & ROE

	Model 6 ROA	Model 6 ROE	Model 7 ROA	Model 7 ROE	Model 8 ROA	Model 8 ROE
(Intercept)	40.78***			−18.77***		
	(1.14)			(2.10)		
ROA	15.84**		−7.41**		−5.19	
	(5.95)		(2.85)		(2.69)	
D/E	0.01	0.01	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ROE		0.56		0.02		0.07
		(0.33)		(0.14)		(0.11)
Size			7.35***	7.21***	5.85***	5.74***
			(0.28)	(0.26)	(0.32)	(0.29)
Advertising Intensity					−1.24	−0.67
					(1.06)	(1.03)
RRI					0.32***	0.32***
					(0.03)	(0.03)
R^2	0.05	0.00	0.35	0.37	0.38	0.38
Adj. R^2	0.05	−0.00	0.34	0.37	0.38	0.38
Num. obs.	11591	11241	10831	10497	10783	10449
Model	Random Effect	Fixed Effect	Fixed Effect	Random Effect	Fixed Effect	Fixed Effect
Robust Standard Error	Yes	Yes	Yes	Yes	Yes	Yes

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$