Håvard Lismoen
Peder Martin Moum

The relationship between ESG and corporate financial performance

An empirical analysis of the S&P 500 and Stoxx 600 companies

Master's thesis in Finance and Investment
Supervisor: Florentina Paraschiv
June 2020

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

NTNU
Norwegian University of Science and Technology

Håvard Lismoen
Peder Martin Moum

The relationship between ESG and corporate financial performance

An empirical analysis of the S&P 500 and Stoxx 600 companies

Master's thesis in Finance and Investment
Supervisor: Florentina Paraschiv
June 2020

Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

NTNU
Norwegian University of
Science and Technology

Abstract

This thesis examines the relationship between corporate financial performance (CFP) and environmental, social and governance (ESG) factors and what implications the relationship has for an investor. The study is based on using aggregated and disaggregated ESG data using Thomson Reuters Asset4. For the corporate financial performance factor, we have focused on profitability and growth. The relationship is explored using companies in the Stoxx 600 and S&P 500 indices. By using a structural equation model (SEM), a panel data fixed effects regression model and a stock market approach using portfolios, we find mixed results for the relationship. By using SEM-models we operationalize environmental, social and governance as latent independent constructs, and growth and profitability are operationalized as latent dependent constructs. The results of the SEM-models indicate that the selected proxies are mostly reliable and have a good fit, but the structural model has very few significant factor loadings which might be caused by omitted variables and/or a poorly specified model.

The panel data fixed effect regression models analyze the relevance of ESG in relationship to annual stock return using data from 2010-2018. The results of the fixed effect regression models indicate that there is a negative relationship between ESG, environmental and governance score towards annual stock return. The social score seems to have a positive, but not a significant effect.

By constructing portfolios based on the ESG score, profitability and low variation in earnings, and pollution (CO2 direct and indirect emissions) we find that companies with the lowest ESG scores (bottom 33%) outperform companies with the highest score (top 33%) in terms of cumulative return using data from 2010-2018. The portfolio based on profitability and low variation in earnings (top 33%) seems to track the return pattern of the bottom 33% ESG portfolio. The portfolio based on 33% lowest CO2 emissions performs better compared to a portfolio of top 33% CO2 emissions. These findings are interesting and contradict some empirical literature that find a positive relationship between ESG and stock market performance.

Sammendrag

Denne studien undersøker relasjonen mellom finansiell prestasjon og ESG for selskap som inkluderes i S&P 500 og Stoxx 600 indeksene, og hvilke implikasjoner denne relasjonen gir investorer. Studien baserer seg på ESG data på et overordnet nivå og et dekomponert nivå, hvor dekomponert data er variabler som inngår i en overordnet ESG score. I forhold til finansiell prestasjon har vi hovedsakelig valgt å fokusere på vekst og lønnsomhet. For å undersøke relasjonen har vi laget ulike SEM-modeller, paneldata regresjonsmodeller og konstruert aksjeporteføljer basert på denne relasjonen. Ved å studere relasjonen på ulike nivå gir det oss et nyansert blikk på sammenhengen, og vi finner ulike resultater. Ved å bruke SEM-modeller forsøker vi å operasjonalisere de latente faktorene hvor faktorer tilknyttet miljø, sosiale forhold og styring er latente uavhengige variabler og vekst og lønnsomhet er latente avhengige variabler. Resultatene av de ulike SEM-modellene viser hovedsakelig at de ulike indikatorene er pålitelige og modellene tilpasser seg data, men strukturmodellen har få signifikante variabler. Dette kan skyldes utelatte variabler og/eller at modellen kan være feilspesifisert. En annen mulighet kan være at sammenhengen er vanskelig å modellere gitt tilgjengelig data.

Paneldata regresjonsmodellene analyserer relevansen av ESG som forklaringsvariabel for å forklare årlig aksjeavkastning ved å bruke data fra 2010-2018. Resultatene fra paneldata regresjonsmodellene antyder at det er en negativ sammenheng mellom score tilknyttet ESG, miljø og styring mot aksjeavkastning. Den sosiale scoren har en antydning til å ha en positiv sammenheng, men ikke signifikant.

Ved å konstruere ulike porteføljer basert på ESG score, lønnsomhet og lav variasjon i årsresultat, og direkte og indirekte CO2-utslipp finner vi at selskap med lav ESG score (laveste 33%) gjør det bedre enn selskap med høy ESG score (topp 33%) i forhold til kumulativ avkastning ved å bruke data fra 2010-2018. Porteføljen med lavest CO2 utslipp gjør det bedre enn den med mest utslipp. Et interessant funn er at porteføljen basert på topp 33% i forhold til lønnsomhet og lav variasjon i årsresultatet har en tendens til å følge avkastningen på porteføljen med lavest ESG score. Et annet interessant funn er at denne studien får motsatte resultater sammenlignet med noe av litteraturen på dette område som har funnet en positiv sammenheng mellom ESG og avkastning.

Preface

This master thesis is the final product of a master`s degree in Economics and Business Administration in the field of Finance and Investment at NTNU Business school. The thesis is written during the spring semester 2020 and is awarded with 30 credit points. This work has given us valuable insight into the relationship between ESG and corporate financial performance. The thesis has allowed us to dive deeper into the topics of sustainability and corporate financial performance, which we are very passionate about.

We would like to thank our supervisor Florentina Paraschiv for helping us in the process and providing constructive feedback and valuable insight. We would also like to thank Michael Schuerle for trying to help us with getting more relevant data for our study.

The authors take full responsibility for the content of this thesis.

Trondheim 10.06.2020

Håvard Lismoen

Peder Martin Moum

Contents

List of figures

List of tables

1 Introduction

1.1 Introduction

The relationship between corporate financial performance (CFP) and environmental, social and governance factors (ESG) is complex. In recent years investors have demanded ESG-data and some have incorporated this data into their decision making and the data has been used in different ways, but it has also contributed to creating a long-term mindset (Eccles, Kastrapeli and Potter, 2017). Climate risk and environmental exposure have also gained attention among investors in both debt and equity markets (Hvidkjær, 2017; Norges Bank, 2019; Ehlers & Packer, 2017). Corporate governance has also received increased attention among financial institutions, for example Goldman Sachs recent statement (Green, 2020) where they refuse IPOs if all their board members are straight, white, males.

Many different strategies and approaches to ESG have been employed in practice and the empirical literature. Some have excluded “sin stocks”, used screening based on environmental, social and governance factors, and others have analyzed the relationship between stock returns and ESG-rating (Hong & Kacperczyk, 2009; Hoepner & Zeume, 2014; Fabozzi, Ma & Oliphant, 2008; Borgers, Derwall, Koedijk & ter Horst, 2013). Hoepner and Schopohl (2018) analyze violations of international norms related to human rights, labor rights and production of controversial weapons in relationship to companies excluded by the Norwegian Government Pension Fund Global, and finds that this exclusion does not financially impair the fund.

There is no universally accepted theoretical framework or definition for ESG, but data providers have been creating latent constructs based on indicators they can measure using corporate disclosure (MSCI, 2020a). There is no current audit standard for this reported data, and audit firms are still in the early stages of developing the ability to audit this kind of data (Eccles, Ioannou & Serafeim, 2014). This leads to some uncertainty on the data reported, but even so, the latent ESG factors are constructed for many companies worldwide by data companies.

This raises the question of how these latent unobservable constructs relate to corporate financial performance for large/mid/small cap companies. How can these proxies be operationalized and divided into subfactors? Does ESG have any explanatory power in relationship to corporate financial performance?

1.2 Research question

The research problem will narrow down the focus of this study. The ambition of this study is to analyze the relationship between ESG and corporate financial performance on an overall level (aggregated data the scores consist of) and using disaggregated ESG factors. The ESG scores are a function of the disaggregated data. More specifically, the goal is to analyze the relationship between environmental, social and governance factors and what proxies these latent constructs consist of. The study will gather data from companies that are included in the S&P 500 and Stoxx 600 indices because large/mid cap companies tend to report more ESG data compared to small firms. To analyze the relationship between environmental, social, and governance factors and corporate financial performance we will construct a Structural equation model (SEM). The purpose of using a SEM-model is to operationalize the latent factors and analyze the relationship between them. For the aggregated data we will use a panel data regression model and Granger causality tests. The research questions can be formulated in this manner:

How does the disaggregated and overall environmental, social and governance factors relate to corporate financial performance? How can the factors be operationalized into latent constructs, and what implications does this have for an investor?

The study tries to analyze the relationship between environmental, social and governance factors and corporate financial performance, and gives implications and insight to investors about these latent constructs. The research questions may give insight for decision making.

This study does not try to generalize the relationship between corporate financial performance and ESG. Corporate policies and actions may change over time given the increased awareness in the recent years, but it is not given that this awareness translates into action. The purpose is to try analyzing the relationship, given observable data, and operationalize corporate financial performance as a dependent latent construct (for the disaggregated ESG data approach). For the overall data approach (aggregated ESG data), the purpose is to analyze the relevance of ESG.

1.3 Overview of the study

In chapter 2 we will present theory and empirical literature that is relevant to analyze the research question. We have focused on using theory explaining investor behavior and the market efficiency. The empirical literature that is presented is based on corporate financial performance, ESG, anomalies and theoretical considerations to ESG investing.

In chapter 3 we will discuss research methods and design. In order to analyze the research question, we have chosen different methodical approaches. The methodical approach includes a cross sectional SEM-model, panel data regression model and a portfolio approach. These approaches will be explained, and we will discuss the reliability and validity of these research methods.

In chapter 4 we will present the results of the tests. In chapter 4.4 and 4.5 we will present SEM-model results. In chapter 4.6 and 4.7 we will discuss the results of the panel data regression and portfolio approach. In chapter 4.8 we will analyze all the results and discuss what implications they have for an investor.

In chapter 5 we will present a summary and a conclusion of the study based on the hypothesis we have presented. The research questions will be concluded based on our findings. Further, we will present suggestions for further research.

2 Theory

2.1 Introduction

In this chapter we will be presenting the relevant literature to shed light on the research questions. The chapters are split into main themes of studies and are used as a backdrop to understand the analysis of the research questions. The purpose of the theory and literature presented here is to get a better understanding of market behavior, what CFP and ESG consist of and the relationship between them, and how ESG-CFP have been studied in the empirical literature. The empirical ESG-CFP literature presented shows implications previous studies have found.

2.2 Efficient market hypothesis

The efficient market hypothesis assumes that prices of securities fully reflect all available information (Bodie et al. 2014). Investors who buy securities in an efficient market should obtain an equilibrium rate of return. Malkiel and Fama (1970) argue that information efficiency is important so investors and companies can allocate their resources in an optimal way. When the market price reflects the fundamental underlying value, the most profitable projects will be prioritized.

Jensen (1978) argues that in an efficient market no one can achieve a return higher than what is expected by the market equilibrium. This implies that every test of the efficient market hypothesis must use an equilibrium model that defines “normal” return. Findings that show that an investor could achieve abnormal returns, could indicate that the market is inefficient, or the theoretical equilibrium model is not correctly specified. Market efficiency is not testable but must be included in a test of the equilibrium model (Fama, 1991).

The efficient market hypothesis comes in different forms and differ by their notions of what is meant by the term “all available information” (Bodie et al. 2014). A weak-form hypothesis asserts that stock prices already reflect all information that can be derived by examining market trading data, e.g. historical prices, trading volume and short interest. In other words, it implies that if data over time would give reliable signals about future performance, all investors would have exploited this signal. A semi-strong-form hypothesis states that all publicly available information regarding the company prospect (e.g. past prices and fundamental data) would be reflected in the price. A strong-form hypothesis would state that stock prices reflect all information regarding the company including insider information.

2.3 Behavioral explanations

Herbert A. Simon (1978) was one of the first to challenge the neo-classical rational assumption. He introduced the term “bounded rationality” which depart from the assumption of “perfect rationality”. Bounded rationality assumes that people “satisfy” rather than “optimize”, and that we make decisions that are rational, but within the limits of the information available. People do not only decide based on calculated self-interest, but for other reasons as well. This study, among others, laid the foundations for behavioral finance today.

Riccardi and Simon (2000) define behavioral finance as attempts to explain and increase understanding of the reasoning patterns of investors, including the emotional process involved and the degree to which they influence the decision-making process. In other words, it attempts to explain what, why and how in relationship to financing and investing from a human perspective. For instance, behavioral finance studies financial markets, anomalies, speculative market bubbles and stock market crashes. Statman (1995) argues that behavior and psychology influence individual investors and portfolio managers’ decision-making process in terms of risk assessment and issues of framing. This can be seen in the process of establishing information of suitable level of risk, and the way investors process information and make decisions depending on how it is presented. There are several definitions of behavioral finance and different understandings of what it consists of. Barber and Odean (1999) argue that behavioral finance enriches economic understanding by incorporating the aspects of human nature into financial modelling. Olsen (1998) describes it as an attempt to comprehend and forecast systematic behavior in order to make correct investments decisions.

Leon Festinger (1957) developed the theory of cognitive dissonance. The theory states that people feel internal tension and anxiety when subjected to conflicting beliefs. As individuals we attempt to reduce our inner conflict in one or two ways (Morton, 1993). We may change our past values, feelings or opinions, and we attempt to justify or rationalize our choice. This theory may apply to investors in the stock markets who attempt to rationalize contradictory behavior so that they seem to follow naturally from personal values or viewpoints. An example of cognitive dissonance is change in our investment beliefs to support our financial decisions. In the 1990s many investors bought internet company stocks without using traditional (fundamental) investment style because the companies had no financial track record. The investors rationalized the change in their investment beliefs by arguing it is a "new economy" and bought stocks simply based on price momentum (Riccardi and Simon, 2000).

Tversky and Kahneman (1974) introduced the term "anchoring" and relate to how an individual creates different points of references for comparison. They argue that people make estimates by starting from an initial value that is adjusted to yield the final answer. They also found that arbitrary numbers could lead participants to make incorrect estimates. By doing different experiments every participant used the initial number as their anchor point. Kahneman (2011) argues that there is no systematic approach to how individuals create an anchor point. In relationship to finance and stock pricing we can assume that investors have an anchor point for "normal" price levels. When many investors see the stock market as "cheap", in comparison to the anchor point, they could make the prices go up.

2.4 The adaptive market hypothesis

The adaptive market hypothesis was introduced by Lo (2004) and is a theory that combines the theory of the efficient market hypothesis with several theories of behavioral finance. By building on Simon's (1987) notion of satisfying, he argues that individuals adapt to a changing environment via simple heuristics. The adaptative market hypothesis uses the conflicting theories of the efficient market hypothesis and behavioral finance to explain investor and market behavior. The theory assumes that people are motivated by self-interest, they naturally make mistakes and they adapt and learn from their mistakes. He argues that rationality and irrationality coexist. The theory believes that people are mostly rational but can become irrational due to high market volatility. Furthermore, the theory argues that investor behavior such as overconfidence, overreaction and risk aversion are consistent with evolutionary models of human behavior. By learning from their mistakes, people will adapt based on failure or success of their strategy.

2.5 Literature review

2.5.1 Market anomalies

Anomalies can be defined as patterns of returns that seem to contradict the efficient market hypothesis and are not predictable by asset pricing models. One example of a market anomaly is Basu's (1983) portfolio study using P/E ratios, and it shows that portfolios of low P/E ratio provided higher returns than high P/E portfolios for the given sample. The P/E ratio effect holds even if returns are adjusted for beta. P/E ratio can be an additional risk indicator and associated with abnormal returns if CAPM is used to establish the benchmark.

Ball and Brown (1968) found another anomaly, namely post-earnings-announcement price drift. This anomaly shows that the stock's cumulative abnormal returns tend to drift for several weeks following a positive earnings announcement. Earnings surprise could be described in many ways, e.g. higher earnings than the average of the analysts. One explanation for this anomaly could be investors' under-reaction to earnings announcements. Another explanation could be a strong connection between earnings and price momentum.

Sloan (1996) found another accounting related anomaly, and his study investigated whether stock prices reflect information about future earnings contained in the accrual and cash flow components of current earnings. By taking a long position in a portfolio with low accruals (high cash component % of net income) and short a portfolio with high accrual (low cash component % of net income) it results in an abnormal return for the given sample period. These examples of anomalies are just a few of many found in the empirical literature.

2.5.2 Corporate financial performance

As noted by Endrikat et al. (2014), corporate financial performance is a multidimensional construct and several classifications have also been introduced for different measures. The most widely used indicator for CFP has been accounting-based performance. Combs et al.

(2005) provide a three-dimensional framework for CFP that includes accounting performance, stock market performance, and growth. Fabrigar, Wegener, MacCallum and Strahan, (1999) point out that one of the dangers of choosing inadequate factors for determining a latent factor is the emergence of spurious connections, and that true connections are obscured. By using well known indicators for CFP, the risk of obscuring true connections is hopefully minimized.

Hamann et al. (2013) advocate the use of four distinct dimensions of performance for firms. These are liquidity, profitability, growth and stock market performance. They argue that these dimensions should be held separated by using factors for performance distinct for each dimension.

2.5.3 Environmental, social and governance (ESG)

It is difficult to distinguish ESG from corporate social responsibility (CSR) because of subjectivity in how one should define it and the terms being closely related. Bowen (1953) was one of the first trying to define what a "socially responsible businessman" is. He argues that corporate social responsibility (CSR) expresses a fundamental morality in the way a company behaves toward society. He further created the foundation by which business executives and academics could consider strategic planning and management decision-making. Carroll (1999) conducts a study of how corporate social responsibility has been defined in the literature going back to the 1960s. He finds that the term evolved into other variants of CSR, such as stakeholder theory (Freeman, 1984) and business ethics theory (Rawls, 1971). However, concepts like corporate social responsibility, sustainability, corporate citizenship (Carroll, 1998), the so-called triple bottom line (Elkington, 1999), or stakeholder management (Freeman & Reed, 1983) were concepts coined not by moral philosophers, but by consultants, activists, or corporate public- relations departments (Norman, 2013).

However, not everyone supported the foundation of CSR. Friedman (1970) argued that a firm's objective is to pursue shareholder value and to maximize financial performance for its shareholders. Jensen (2002) and Tirole and Bénabou (2010) also support this statement and argue that social responsibility diverts from maximizing financial performance because CSR comes with a cost, therefore making it a disadvantage.

Van Marrewijk (2003) argues that there is no point of trying to define an all-inclusive definition of CSR and corporate sustainability. He argues that the “all -inclusive” definition should be abandoned, and various specific definitions should be accepted. Krüger (2015) argues that CSR has different interpretations for different stakeholders, and that it also implies the social and environmental dimensions, while ESG has an additional governance dimension. Stellner et al. (2015) argues that there is no universally accepted definition of CSR, and the environmental, social and governance dimensions should be included in the definition.

The term environmental, social and governance (ESG) factors in relationship to finance goes back to 2004 and is a result of cooperation between the finance industry and UN Global Compact that created a report titled “Who Cares Wins” (UN Global Compact, 2004). The purpose of the cooperation was to address and integrate ESG issues in asset management, securities brokerage services and research. This resulted in implementing universal principles in business by establishing a link between the ESG issues and investment decisions related to these factors. The awareness of ESG factors existed long before this report, but no unified global framework existed due to the complexity. The report argues that an economy is dependent on a healthy civil society which is dependent on a sustainable planet. Therefore, investment decisions should have a clear self-interest in contributing to better management of social and environmental impacts. By taking ESG factors in consideration, the report argues that it may contribute to more stable and predictable financial markets because of transparency.

Before this report, the financial analysts had issues defining ESG and measuring the business case. Another problem was quality and quantity of information and the analysts short-term focus e.g. quarterly. The report also operationalizes ESG into measurable variables and sub-factors. Companies implementing these factors may increase share value by managing risks related to emerging ESG issues by anticipating regulatory changes, consumer trends and accessing new markets or reduce costs. A survey conducted among European fund managers, analysts and investor relations officers found that 78% believe that environmental and social risk have a positive impact on a company's long-term market value (UN Global Compact, 2004).

Despite this report, it is only in recent years that the awareness of ESG investing has increased in the stock and bond markets (Ehlers & Packer, 2017). The market for green bond issuance has increased from 2 billion USD in 2010 to 60 billion USD in 2017. The world's largest asset management company, BlackRock, expects the global ESG exchange traded fund market (ETF) to be around 400 billion USD in 2028 (Blackrock, 2018). In the same period, 2010-2017, the Social Responsible Investment world index (SRI) has been doubled. The SRI is based on ESG data and exclusion of companies which have negative social or environmental impact (MSCI, 2020b). Several other indices variants have been created in recent years, and the purpose is to take climate change risk, social inequality, governance and transparency in consideration.

However, ESG as a measure has been heavily criticized by Porter, Serafeim and Kramer (2019). They argue that ESG score is a myriad of metrics with little consideration of their financial materiality. Furthermore, they argue that these ESG criteria have been developed without regard to the causal link between the company's social impact and its bottom line. Even though ESG reporting has become more detailed in recent years, they argue that another problem with the ESG score is that the companies are judged on their overall performance, equally weighted, rather than the most salient issues of their businesses.

2.5.4 Theoretical considerations to ESG investing

For an investor who does not have inside information about firm values and does not engage in active ownership to assert influence over the management, the central question is not whether ESG initiatives by firms create value, but whether any such value is properly recognized by the stock market (Hvidkjær, 2017). He argues that underreaction to ESG information is the main argument for outperformance, and the value of positive ESG effects is not recognized by the stock market. Further he states that this is a plausible hypothesis, given evidence exists that the stock market underreacts in various situations. For example, post earnings drift announcement (Ball and Brown, 1968) and momentum (Jegadeesh & Titman, 1993) are evidence against market efficiency and underreaction may exist. Another argument is the valuation of intangible assets and underreaction. Edmans (2011) argues that there is evidence of underreaction to intangible assets such as R&D and likewise for ESG investments. ESG investments are usually intangible as well, but also tangible.

Hvidkjær (2017) also argues that another reason for outperformance is that ESG investing has become more popular over time. The growing demand for "ESG-stocks" may push the price up, especially in markets where there are few ESG investment opportunities. In other words, the demand effect may affect the valuation. Merton (1987) argues that when a large group of investors ignore certain stocks, they may become undervalued. The question is how this may affect high/low ESG score stocks. Given that the undervaluation is "permanent", a permanent low price implies higher dividend/price ratio and higher return, all else equal. This will also affect the sin-stocks and may imply lower returns.

From a diversification perspective based on Markowitz (1959), exclusion of entire industries or sectors may affect broad portfolio risk-return trade-off. The question is how this will affect the optimal risk-return trade-off. If ESG information does not affect pricing, there is no point in exclusion based on the risk-return relationship and vice versa if it does. In other words, ESG restrictions may or may not affect the optimal portfolio. Another important factor is the cost of ESG information and screening, which is crucial for passive low-cost investment strategies. A lot of ESG data are available and reported in databases such as Thomson Reuters or Bloomberg, but the licenses may be very expensive for an individual investor. Some data are also available in companies annual (or quarterly) report. Obtaining this information may be very challenging as an individual, especially when it involves picking individual stocks.

Furthermore, we must consider ESG investing penetration in the long run. Given a high level of awareness and penetration of ESG related investing, it is hard to see how outperformance could be sustained (Hvidkjær, 2017). The effect of underreaction of ESG information may disappear if many investors pursue such a strategy, the demand may be temporary and ignored stocks may become more relevant. The question is whether a large portion of investors pursuing ESG strategies causes underperformance. Of course, this is not given and Hvidkjær (2017) argues that we must take it into consideration how close we are to a steady-state level of ESG investing and Merton's (1987) argument of ignored stocks.

It is important to note that we are dealing with complex terms that are comprised of multiple different factors. This is true for the combined expression for ESG, as well as the individual E, S and G terms. Endrikat et al. (2014) mentions that there exists no commonly shared understanding of the term environmental performance, and that different studies use different measures for environmental factors.

The social dimension is also complex in nature, as noted by Devinney (2009) where he points out that the science of CSR is suffering because there are so many different aspects encompassing this term that are trying to combine it all will not produce any empirical rebuttal or validation. Love (2011) states that a source of bias in her meta study comes from the fact that so many ways are used to operationalize the governance factor across different studies.

2.5.5 Empirical ESG and corporate financial performance literature (CFP)

A fundamental question in the ESG-CFP literature is how the ESG factors affect an investor's portfolio and the risk-return characteristics of the portfolio (Hvidkjær, 2017). Previous literature has looked at "sin-stocks" relative to various benchmarks, ESG ratings and screening in relationship to returns, event studies that indicate that the stock market does not respond positively to ESG initiatives by firms, ESG in relationship to the cost of capital and how active ownership in relation to ESG can create value for shareholders and stakeholders (Hvidkjær, 2017).

Hong and Kacperczyk (2009) investigate the effect of negative screening for sin-stocks defined as U.S tobacco, alcohol and gambling firms. These stocks are neglected by many institutional investors. They found that sin-stocks outperform comparable stocks by 3-4% return yearly using 1926-2006 as a sample, but not all results are robust controlling for analyst coverage and market-to-book value as a control variable. The returns are calculated using a Fama-French factor model, but only significant at the 10% level for the standard 3-factor model.

Kempf and Osthoff (2007) construct long-short value-weighted portfolios from the S&P500 and DS 400 stocks in the period 1992-2004. They find 4-factor significant alphas of around 5% year using data from 1992-2004 using industry-adjusted ESG scores. Borgers, Derwall, Koedijk and Horst (2013) show that the ESG outperformance in Kempf and Osthoff's study is significant until 2004, and after that they are close to zero and insignificant. This goes to show that the time aspect can have an impact on the effect of ESG.

Auer (2016) studies the effect of exclusionary screening on portfolio Sharpe ratios using ESG ratings for the companies included in the Stoxx 600 index using 2004-2012 data. The main result of the study is that the Sharpe ratio of the stocks increases when excluding stocks with poor governance rating, while exclusionary screening based on environment and social factors does not affect Sharpe ratios. However, the sample period is short, so the test power is low.

Some studies seeking to investigate the relationship between ESG and CFP have focused on a specific geographical area. Velte (2017) uses regression on data from companies based in Germany, and finds a significant positive connection between ESG score, individual pillar score and CFP represented by ROA, but no significant result for a connection to Tobin's Q. Doque-Grisales and Aguilera-Caracuel (2019) look at the connection between ESG and CFP for multinational companies operating in emerging markets based in South-America, where they find a negative connection. Hoang, Przychodzen, Przychodzen and Segbotangni (2020) use disaggregated environmental factors in a regression analysis of data collected from 361 U.S companies and find that greenhouse gas emissions generally seem to be the most influential environmental factor towards CFP. However, the connection seems to differ between both negative and positive considering what financial measure is used.

In a second level meta study done by Friede, Busch and Bassen (2015) the point of interest is specifically the relationship between ESG and CFP. They conducted their meta study with basis on 60 other meta studies concerning this subject. They find that for equities, the existing research shows 52,2% positive relationship between ESG and CFP, while 4,4% are negative. In non-portfolio studies, a total of n=568, they find a positive connection in 56,7% of the studies, while 5,8% are negative. Neutral or mixed results comprise the last 37,5%. Their results show an overall positive connection between ESG and CFP. However, it must be noted that this meta study is from 2015, and several later studies have investigated the subject. Therefore, the total percentage in this line of research may have changed.

Khan, Serafeim and Yoon (2016) analyze the relationship between CFP and ESG by classifying ESG data as material and immaterial on an industry level. By creating stock portfolio return regressions and firm level panel regressions, they find that companies with good ratings on material sustainability significantly outperform companies with poor ratings. They also find that companies with good ratings on immaterial sustainability do not significantly outperform companies with poor ratings.

The large body of literature concerning the relationship between ESG and CFP finds different results, much depending on which ESG measures they incorporate and what financial performance factors are included. The differing results also highlights the complexity of the term ESG and the lack of a set standard in both reporting and database use. The literature also provides some insight into what implications the relationship has for investors.

2.5.6 Environmental screens

Guenster, Bauer, Derwall and Koedijk (2011) use Innovest eco-efficiency data with measures on operating performance and equity valuation. They find that eco-efficient companies become more expensive, as measured by Tobin's Q, from 1997 to 2004. Halbritter and Dorfleitner (2015) used a long-short 4-factor model approach yielding an alpha of 6,6% per year during the sample period 1990-2001. For the sample period of 2002-2012 they find insignificant and negative alphas. Statman and Glushkov (2009) found no evidence of outperformance based on KLD environmental scores from 1992-2007.

2.5.7 Social screens

Edmans (2011) explores the relationship between employee satisfaction and stock returns. He found that a value-weighted portfolio of the "100 Best Companies to Work for in America" earned an annual four-factor alpha of 3,5% from 1984 to 2009. The model controls for firm specific characteristics and different weighting methodologies. Edmans (2011) argues that the market fails to incorporate the intangible information, and the prices are corrected as the information become tangible through higher earnings.

2.5.8 Governance screens

Gompers, Ishii and Metrick (2003) construct a firm-level governance index over shareholder rights. A firm with weak shareholder rights would have a high index score and strong governance would have a low index score. They use a sample of 1500 large US firms from 1990-1999, and they create portfolio that is long in the 10% lowest scoring and 10% short in the highest scoring companies. The portfolio yielded an abnormal return of 8,5% per year. Bebchuk, Cohen and Wang (2013) extended the sample size of Gompers, Ishii and Metrick (2003) to cover 1990-2008. They found that the abnormal returns are insignificant during 2000-2008. They also argue that "good governance" firms tend to report more positive earnings surprises than poor governance firms in the 1990s, but the relationship disappears in the 2000s.

Gu and Hackbarth (2013) use Gompers, Ishii and Metrick (2003) as a base, and identifies that the relationship between stock returns and governance is concentrated among high transparency firms (as measured by the distribution of analyst's forecasts). They argue that highly transparent firms are more valuable takeover targets because acquirers can bid more effectively and identify synergies more precisely.

2.6 Conceptual SEM-model

The SEM-model is often used to find a connection between observable indicators and latent factors. We have used an explorative approach for the ESG and CFP indicators because we have no benchmark model and the literature is lacking. See 3.2 for description of the SEM-models and 4.4-4.5 for results.

The goal of a SEM model is to understand the pattern of correlations between different variables and explain as much of the variance as possible with a research model specified (Bowen & Guo, 2012). Before constructing an empirical model, it is important to have an already established idea for a scientific model which is based on prior research or empirical studies (Bowen & Guo, 2012). Below is the conceptual model for the latent variables we would like to test.

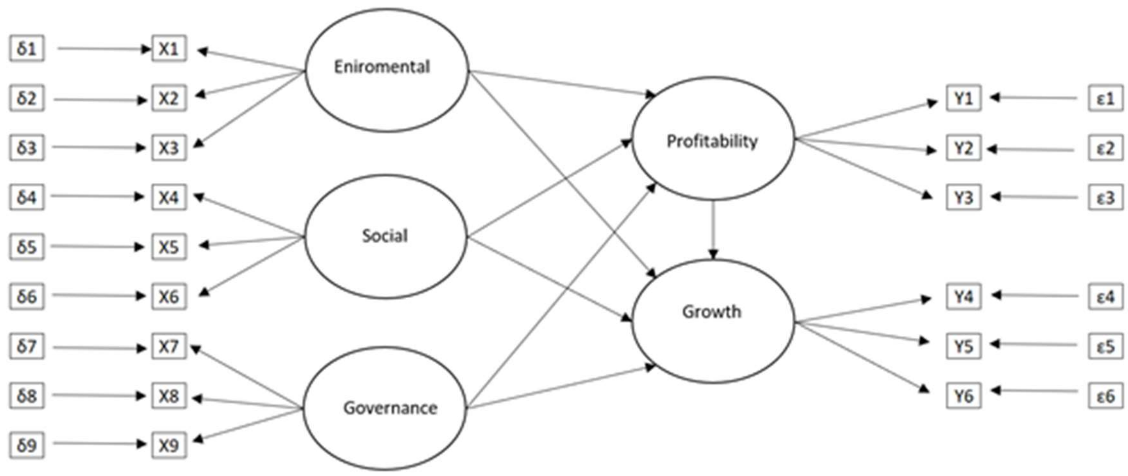


Figure 2.1 Conceptual research model using structural equation models (SEM)

Based on the literature presented in chapter 2, this conceptual model became the basis for our further work with the concept of ESG and CFP. There is little unity among investors and financial institutions about what ESG should consist of, and we wanted to expand the knowledge of the subject. This is done by seeing what independent variables combine into different factors by setting up an explorative factor analysis. Further, the factors will be tested by using a confirmatory factor analysis and looking for Granger causality for some ESG variables. The variables used as indicators for the environmental, social and governance factors is based on MSCI (2020a). The variables are then combined in a full SEM-model and analyzed using LISREL 10.0.

The main purpose here will be to expand the term ESG and shedding light on factors used by Thomson Reuters Asset4 database. This can be done by investigating whether these factors can contribute to some explanatory power in relationship to the corporate financial performance of different companies. The main difficulties of this are that the terms in the model are complex and the data reported by different companies will vary.

2.6.1 Operationalization of the latent variables

The complex conceptual model emphasizes the importance of operationalization of the latent constructs to ensure term validity. Due to lack of data and poor quality for the social and governance factors, many latent factors are impossible to operationalize using Thomson Reuters Asset4. The latent factors that are possible to test will somewhat be linked to core operations. The operationalization will be based on MSCI (2020a) using an explorative approach and test different models. The goal is not to operationalize the "entire" environmental, social and governance dimensions, but different subfactors given what data are available. The hypothesized latent constructs will be shown in chapter 4.

Environment

The environmental variables used are based on resource use and pollution for the different companies. All the reported corporate environmental data regarding pollution are estimates, so measurement error could occur.

Social

Some of the social data that is available is hard to separate into factors due to being closely related. The focus here is to operationalize social policies that may affect core operations. However, the data that are available has mixed quality in terms of richness of information.

Governance

The focus using latent governance factors is taking polices regarding management and corporate behavior into consideration. The challenge in operationalizing these factors is also that the data available are hard to separate into unique factors. Almost all governance data that are available are dichotomous variables.

Corporate financial performance

The operationalization of CFP is based on Hamann et al. (2013). CFP is separated into profitability, liquidity, growth and stock market performance. They have tested different variables and found indicators with good fit. We will mainly focus on operationalizing profitability using indicators based on NBIM (2015) which are ROA, ROE and ROIC. Growth will be operationalized as 1-year employee growth, 1-year total asset growth and 1-year net sales growth. Other indicators will also be tested.

2.6.2 SEM-model hypotheses

Based on the literature in chapter 2 and the developed conceptual model, four hypotheses have been developed.

H1S: The latent environmental factor has either a significant positive effect or a significant negative effect on profitability and growth.

This view is derived from different sources. Gallego-Alvarez, Segura and Martinéz-Ferrero (2014) find that environmentally friendly policies are positive for corporate financial performance. The findings from Lewandowski (2017) suggest that making progress towards mitigating climate change has a negative effect on stock prices, while Busch and Hoffmann (2011) find mixed results. Busch and Hoffmann (2011) find a positive connection between lower greenhouse gas emissions and return for investors. They found a negative effect for the connection between the way companies address climates change and accounting-based CFP.

H2S: The latent social factor has a positive contribution to profitability and growth. Companies with socially responsible practices will have an overall better corporate financial performance than those who do not.

Benson and Davidson (2010) find that firms with a higher aggregate stakeholder management scores have a higher firm value. Crook, Ketchen, Combs and Todd (2008) find a significant positive relationship between socially responsible practices and corporate financial performance in their meta study.

H3S: The latent governance factor has a positive or negative significant effect on profitability and growth.

When it comes to governance, the literature has some mixed results as well. For example, Lai, Li and Li (2016) find no result significantly different from zero for portfolios differentiated on governance factors. This is also supported by Core, Guay and Rusticus (2006), who find no significant results for the governance factor. Gompers, Ishii and Metrick (2003) find that firms with weak shareholder rights exhibit significant stock market underperformance. Chen et al. (2007), find that firms with good corporate governance outperform those with weak corporate governance.

H4S: Profitability has a significant positive or negative effect on growth.

This is supported by Cho and Pucik (2005) where they find a significant positive effect for growth on profitability. Ramezani, Soenen and Jung (2002) found that their measures of corporate profitability and value for shareholders generally rise with growth, but at a certain level of growth it adversely affects profitability. The firms that exhibit moderate growth generally have a higher value creation for their owners. Since we investigate the relationship from profitability to growth, we do not set any condition on the direction of the effect.

2.7 Hypotheses panel data regression models

Below we present four different hypotheses used to investigate the relationship in the panel data regression models. The hypotheses are also based on the literature in chapter 2. These hypotheses are named with a P to the end of it, so it is not confused with the hypotheses for the SEM-models.

H1P: ESG score has a significant positive or negative effect on yearly stock returns.

Whether ESG score has a positive or negative effect on stock returns is not clear in the empirical literature, and it is not given that a high ESG score from one year to another impacts the return. However, given recent investor attention it might be unclear what the effect is. There is also evidence for investors reacting negatively to positive CSR news (Krüger, 2015). Franzén (2019) also analyzes the effect of ESG, environmental, social and governance factors on stock returns and finds mixed results in terms of positive/negative signs and level of significance. Their analysis is based on companies in the S&P 500 going back to 2002 and concludes that there is no reliable evidence that ESG and its pillars (E, S and G) have any significant explanatory power on stock returns.

H2P: ROA and ROE have a positive significant effect on yearly stock returns.

Another question is if fundamental indicators of corporate financial performance are good indicators for stock returns. These variables are indicators of quality, and we would assume that companies with a high ROA and ROE over time would yield a positive return. However, the effect might be unclear for the large sample size. This has been pointed out by NBIM (2015), where they state that the quality factor, where ROA and ROE are included, has a positive effect on stocks when used in a portfolio setting.

H3P: Lagged (1) ROA and ROE have a significant positive effect on yearly stock returns.

Using the lagged values of ROA and ROE may have more predictive power based on the assumption that the stock markets react to earnings surprises in the annual income statement and the quality anomaly. However, on a yearly basis, other factors and external events or happenings can affect the return. The quarterly reports may also contain this information and make financial information in the annual income statement less relevant.

H4P: Environmental, social and governance pillar scores have a positive or negative significant effect on yearly stock returns, and controversies score has a positive significant effect on yearly stock returns.

There is no clear evidence that having a high environmental, social and governance score is rewarded in the stock market (Franzén, 2019). However, we would assume that over time less environmentally friendly companies which e.g. are very CO_2 intensive, might be forced to choose more environmentally friendly solutions which might come at a significant cost, but might not be the case for the given sample. Companies who are lacking in reporting are also penalized with a lower score. The social and governance effects might have a more indirect effect on the core operations, and the signs could be plus or minus. Aouadi and Marsat (2018) analyzed the relationship between market value and ESG controversies score using a sample of 4000 companies from 58 countries. They found that the ESG controversies score is positive and significantly related to stock returns.

3 Research method

3.1 Introduction

In this section the research methods used to analyze the relationship between ESG and CFP, and the validity and reliability for the methodical approaches are discussed. To analyze the relationship, we have used a cross-sectional approach by using SEM-models and a time series approach by using panel data regression and a portfolio approach. The reasoning for this approach is that the relationship is complex and needs to be analyzed using different methodological approaches.

3.1.1 Description of data and sample of companies to be analyzed

Appendix 8 shows the full list of variables we have used to explore the relationship between CFP and ESG. For the environmental factor, the variables based on pollution are continuous and some are dichotomous variables. The social and governance factors are complex and hard to measure, and most of the data are dichotomous variables. A challenge using ESG data is missing data and inconsistency which limits the sample size. Most of the proxies for CFP data are accounting related data.

The S&P500 is an index for large cap U.S companies and covers approximately 80% of U.S market capitalization (S&P Dow Jones Indices, 2020). Stoxx 600 is an index for large, mid, and small cap companies based on 17 European countries (Stoxx, 2020). Both indices are value weighted. The reason for choosing the constituents of these indices is the availability of ESG data which is a prerequisite.

3.2 SEM-models

To investigate the term ESG and the connection between the data contained in the ESG score and corporate financial performance, a SEM model will be used. This is a powerful tool that utilizes the covariance matrix for the data collected and sees whether this can be explained by a model one has specified. This tool excels at finding connections between data if the model specification is good. The purpose of using a SEM model is taking unobservable latent factors in consideration when exploring the relationship between corporate financial performance and ESG, something which cannot be observed directly.

3.2.1 Time period and dataset for ESG in SEM model

The dataset for the SEM model is gathered from Thomson Reuters Datastream. This is a financial database that contains a multitude of data from different financial instruments. In addition to this, Thomson Reuters also contains data for the ESG variables that are used to calculate the ESG score for firms across the globe.

The extracted data consist of all the companies contained in the S&P 500 and Stoxx 600 indices at the start of February 2020. The datapoints included are composed of both ESG measures and financial performance data for all the companies. Ideally one would like to have a longer time dimension than one year, but there is inconsistency in reporting from the companies and many lack datapoints in earlier years. The explorative and confirmatory factor analysis and the SEM models are based on using data from 2018. We will focus on creating a reliable model instead of testing different time periods. See appendix 8 for a list of ESG data and financial performance variables that have been used.

To have a set of data with as many reported datapoints as possible, the chosen companies are comprised of the firms listed in S&P 500 and Stoxx 600 indices. The reason why is that both U.S and European companies report many of these factors compared to other countries. For example, Hassan & Romilly (2018) end up with most companies coming from the US and UK when looking at a global dataset. Research also suggests that ESG factors have an effect for companies in countries outside Europe and the U.S. A study by Utz (2018) compared Japanese, European, US and Asia-Pacific corporations found that the effect of CSR (ESG) was a significant predictor for lower idiosyncratic risk across all the regions, implying that ESG has an effect not only in western countries.

In order to minimize the amount of missing data in the sample, it is necessary to make a qualitative and quantitative assessment of the dataset and decide which ESG factors we can include and which factors are lacking too much data to perform an analysis on. This is accomplished by looking at the different factors and seeing how many of them are missing.

Furthermore, we will conduct an exploratory and confirmative factor analysis to find which factors comprise a latent factor.

There are not many previous studies in this field that use disaggregated ESG factors. Therefore, it is necessary to expand the knowledge concerning the ESG factors, and their connection to financial performance. This is highlighted by Endrikat et al. (2014) which shows the need for more research into the circumstances shaping the link between environmental performance and financial performance. Friede et al. (2015) point out that future research should look at the effect of specific ESG sub-criteria on CFP to expand on the understanding of their possible connection.

3.2.2 Test specification SEM-models

3.2.2.1 Estimation technique

The SEM-models use the maximum likelihood (ML) estimation technique and algorithms to generate starting values. There are several other estimation techniques, but all techniques are dependent on sample size, type of data and distribution. ML fits the sample sizes we are working with, and it is the most used technique and is usually the default estimation technique in SEM statistical software (Bowen & Guo, 2012). The ML estimator is defined as

$$\log L = -\frac{1}{2}(N-1)\left\{\log \sum(\theta) + tr(S\sum(\theta)^{-1})\right\} + c \tag{1}$$

where log is the natural logarithm, L is the likelihood function, N is the sample size, t is the parameter vector, σ-θ is the model implied covariance matrix and | σ-θ | its determinant, tr is the trace matrix and c is a constant that contains terms of the Wishart distribution that do not change once the sample is given (Schermelleh-Engel, Moosbrugger & Müller, 2003).

ML estimation technique also have assumptions and assume that data are continuous and multivariate normal distributed. ML also assumes that σ-θ are positive defined, and the matrices must be nonsingular. Bollen (1989) found that if the model is specified correctly and the sample size is sufficiently large, ML provides parameter estimates and standard errors that are asymptotically unbiased, consistent, and efficient.

A limitation of ML is the assumption of multivariate normality. Violation of this assumption can lead to very misleading results. However, ML seems to be quite robust against violation of the normality assumption (Boomsma & Hoogland, 2001; Chou & Bentler, 1995; Curran, West & Finch, 1996; Muthén & Muthén, 2002; West, Finch & Curran, 1995). Simulations suggest that ML parameter estimates are still consistent, but not necessarily efficient. Satorra and Bentler (1994) developed a correction for ML so it could account for nonnormality which is a robust estimation technique that has good statistical properties. The robust estimation requires an asymptotic covariance matrix that corrects for skewness and kurtosis in addition to the model implied covariance matrix. The nonnormality test developed by Mardia and Foster (1983) shows that skewness, kurtosis and joint skewness and kurtosis can be tested, which follows an approximate Chi-square distribution. For all the following models we will test for nonnormality using the following hypothesis and test statistics:

$$\begin{array}{lll} \textit{Skewness} & \textit{Kurtosis} & \textit{Skewness}\text{ and kurtosis} \\ \mathrm{H}_0: \ \mathrm{M}\gamma_1 = 0 & \mathrm{H}_0: \ \mathrm{M}\gamma_2 = 0 & \mathrm{H}_0: \ \gamma_1 = \gamma_2 = 0 \\ \mathrm{H}_1: \ \mathrm{M}\gamma_1 \neq 0 & \mathrm{H}_1: \ \mathrm{M}\gamma_2 \neq 0 & \mathrm{H}_1: \text{ Both not equal to zero} \end{array} \tag{2}$$

$$Z_k = 3\sqrt{(d/2)} * \left\{ 1 - (2/9d) - \left[\frac{1-(2/d)}{1+e\sqrt{2/(d-4)}} \right]^{1/3} \right\} \tag{3}$$

$$Z_S = \frac{\left\{ (27Nk^2(k+1)^2(k+2)^2 b_{1,k})^{1/3} - 3k(k+1)(k+2) + 4 \right\}}{\sqrt{12k(k+1)(k+2)}} \tag{4}$$

$$C_{sk} = Z_s^2 + Z_k^2 \tag{5}$$

3.2.2.2 SEM goodness of fit indices

There are several goodness of fit indices, but we are using the indices that are most used which is based on Schermelleh-Engel, Moosbrugger and Müller (2003).

Chi-sqaure test

The chi-square tests if the population covariance matrix is equal to the model implied covariance matrix. The hypothesis and the test statistics are the following:

$$\mathrm{H}_0 : \sum = \sum(\theta) \text{ and } \mathrm{H}_1 : \sum \neq \sum(\theta) \qquad (6)$$

$$\chi^2(df) = (N-1)F(S, \sum(\hat{\theta})) \qquad (7)$$

The test has strict statistical assumptions and will often reject the null hypothesis when the sample size is large (Hammervold & Olsson, 2012). The test assumes that the implied model holds in the population. For models that deviate from the multivariate normal distribution we are using the Satorra and Bentler (1988) Chi-square (C3) that corrects for nonnormality. Rejecting the null hypothesis implies that the data does not conform to the model, but we must take all the goodness of fit indices into account.

Root mean square error of approximation (RMSEA) and close-fit-test

RMSEA is a less strict statistical test compared to the Chi-square-test and RMSEA measures the deviation per degree of freedom between the implied covariance matrix and sample covariance matrix. Using the following hypothesis and test statistics we test for close fit.

$$\mathrm{H}_0 : \text{ EA } \leq 0{,}05 \text{ and } \mathrm{H}_1 : \text{ EA} > 0{,}05 \qquad (8)$$

$$\hat{\varepsilon}_a = \sqrt{\max\left\{\left(\frac{F(S, \sum(\hat{\theta})}{df} - \frac{1}{N-1}, \ 0\right)\right\}} \qquad (9)$$

RMSEA as close to zero as possible indicates good fit (Steiger, 1990). Browne and Cudeck (1993) define “close fit” as RMSEA value less than or equal to 0,05. Browne and Cudeck (1993) also argue that values between 0,05 and 0,08 is an adequate fit and values between 0,08 and 0.10 as mediocre fit and values greater than 0,10 as not acceptable fit. Hu and Bentler (1999) suggest that RMSEA of less than 0.05 should be a cutoff-criteria. The close-fit test is a variant of the Chi-square-test but using a non-central distributed chi-square and uses RMSEA or the p-value as a test statistic. A p-value over 0,10 indicates good fit and a value between 0,05 and 0,1 indicate acceptable fit.

Standardized root mean square residual (SRMR)

SRMR is an index for the average of the standardized residuals between the sample covariance matrix and the estimated covariance matrix. The index is dependent on sample size and is sensitive for not correctly specified models (Hu & Bentler, 1998; Schermelleh-Engel et al., 2003). A rule of thumb, based on Schermelleh-Engel et al. (2003), is that values under 0,05 is a good fit and values between 0,10 and 0,05 is an acceptable fit.

The residuals are first divided by the standard deviation $S_i = \sqrt{S_{ii}}$ and $S_j = \sqrt{S_{jj}}$ of the respective manifest variables, which leads to a standardized residual matrix

$$r_{ij} = \hat{\sigma}_{ij} / (S_i S_j) \tag{10}$$

where r_{ij} is the observed correlation between the respective variables.

Goodness of fit index (GFI) and adjusted goodness of fit index (AGFI)

GFI measures the relative amount of variance and covariance in the empirical covariance matrix that is predicted by the model-implied covariance matrix (Jöreskog & Sörbom, 1989). The test implies testing how much better the model fits as compared to “no model at all”, e.g. all parameters fixed to zero (Schermelleh-Engel et al., 2003).

$$GFI = 1 - \frac{F_t}{F_n} = 1 - \frac{\chi_t^2}{\chi_n^2} \tag{11}$$

Where χ_n is the chi-square of the null model, χ_t is the chi-square of the target model and F is corresponding minimum fit function value.

The GFI index ranges between zero and one, where values close to one indicate good fit. The usual rule of thumb for this index is that 0.95 is an indication of good fit relative to the baseline model and 0.90 is an acceptable fit (Marsh & Grayson, 1995; Schumacker & Lomax, 1996).

Jöreskog and Sörbom (1989) developed the adjusted goodness-of-fit index to adjust for bias resulting from model complexity. AGFI adjusts for the model's degrees of freedom relative to the number of observed variables and therefore rewards less complex models with fewer parameters. AGFI is given by:

$$AGFI = 1 - \frac{df_n}{df_t}(1 - GFI) = 1 - \frac{\chi_t^2 / df_t}{\chi_n^2 / df_n} \tag{12}$$

where χ_n is the chi-square of the null model, χ_t is the chi-square of the target model and df_n is the number of degrees of freedom for the null model and df_t is the number of degrees of freedom for the target model. AGFI range between zero and one, with larger values indicating a better fit. A rule of thumb for this index is that 0,90 indicate good fit relative to the baseline model and values greater than 0,85 may be considered as an acceptable fit. Both indices decrease with increasing model complexity, especially for smaller sample sizes (Anderson & Gerbing, 1984).

Normal fit index (NFI) and Nonnormed fit index (NNFI)

NFI ranges from 0 to 1 and a higher value indicates a better fit (Schermelleh-Engel et al., 2003). The usual rule of thumb for NFI is that 0,95 indicate a good fit relative to the baseline model and values greater than 0,90 is an acceptable fit (Marsh & Grayson, 1995; Schumacker & Lomax, 1996).

$$NFI = \frac{\chi_i^2 - \chi_t^2}{\chi_i^2} = 1 - \frac{\chi_t^2}{\chi_i^2} = 1 - \frac{F_t}{F_i} \tag{13}$$

A problem with NFI is that it is affected by sample size (Bearden, Sharma & Teel, 1982). Bentler and Bonnet (1980) developed NNFI to handle this problem. NNFI also ranges from 0 to 1. A rule of thumb for NNFI is that 0,97 is an indicator of good fit relative to the independence model and values greater than 0,95 may be interpreted as an acceptable fit. More complex models (less restrictive) are penalized by a downward adjustment. NNFI is less affected by sample size (Bentler, 1990; Bollen, 1990; Hu & Bentler, 1995, 1998)

$$NNFI = \frac{(\chi_i^2 / df_i) - (\chi_t^2 / df)_t}{(\chi_i^2 / df_i) - 1} = \frac{(F_i / df_i) - (F_t / df_t)}{(F_i / df_i) - 1/(N-1)} \tag{14}$$

Comparative fit index (CFI)

CFI compares the model with an alternative independence model and ranges from 0 to 1 (Schermelleh-Engel et al., 2003). A rule of thumb is that 0,97 indicate good fit relative to the independence model and values greater than 0,95 may indicate an acceptable fit. CFI is less affected by sample size (Bentler, 1990; Bollen 1990; Hu & Bentler, 1995, 1998, 1999).

$$CFI = 1 - \frac{\max\left[(\chi_t^2 - df_t), 0\right]}{\max\left[(\chi_t^2 - df_t), (\chi_i^2 - df_i), 0\right]} \tag{15}$$

Description measures of model parsimony

Parsimony is important in assessing model fit and serves as a criterion for choosing between alternative models (Hu & Bentler, 1995). Parsimony Goodness-of-fit Index (PGFI), Parsimony Normed Fit Index (PNFI) and Akaike Information Criterion (AIC) adjust for model parsimony when assessing the fit of SEM-models.

Parsimony Goodness-of-Fit Index (PGFI), Parsimony Normed Fit Index (PNFI) and Akaike Information Criterion (AIC)

PGFI and PNFI are modifications of GFI and NFI (Mulaik et al., 1989; James et al., 1982). PGFI and PNFI both range between 0 and 1 and higher values indicating more parsimonious fit, but they are not standardized between 0 and 1.

$$PGFI = \frac{df_t}{df_i} GFI \tag{16}$$

$$PNFI = \frac{df_t}{df_i} NFI \tag{17}$$

AIC adjusts the Chi-square for the number of estimated parameters and can be used to compare models. It is not possible to interpret an isolated AIC value, and the minimum AIC value of another comparable model is regarded as the best fitting model.

$$AIC = -2\log L + 2t \tag{18}$$

3.2.3 Panel data regression models

To account for individual heterogeneity and time specific differences we have created different panel data models to explore the relationship between corporate financial performance and ESG for the S&P500 and Stoxx 600 companies. The purpose of these models is to get a better understanding of the data by looking at specific industries, within industry and cross-industry differences. To look at these differences we have used descriptive statistics, analysis of variance and panel data regression. The panel data models are also used to test for Granger-causality, see section 3.5. All the panel data models will be tested for autocorrelation and heteroskedasticity in chapter 4.

3.2.3.1 Time period and dataset for the panel data models

The sample for the models is based on annual data from 2010 to 2018. For a full list of all variables see appendix 8**.** The ESG data used for panel data regression is the ESG scores for the individual companies. By using disaggregated ESG data from the Thomson Reuters database, missing data would be a severe problem and might lead to sample bias. Nevertheless, ESG score values in some of the panel data models are missing due to the fact that not all companies are rated for the entire period. For the corporate financial performance data, we have used proxies based on NBIM (2015) and Hamann et al. (2013).

3.2.3.2 Panel data regression: fixed or random effects

To determine the type of regression model, we used the Hausman-test based on Hausman and Taylor (1981). The test has the following hypothesis and test statistics:

$$\mathrm{H}_0: \text{ Random effets and } \mathrm{H}_1: \text{ Fixed effects} \tag{19}$$

Random effects: $\mathrm{Y}_{it} = \beta_0 + \beta_1 X_{it} + (\alpha_i EF_i + \rho_t ET_t + v_{it}) = \beta_0 + \beta_1 X_{it} + (\varepsilon_{it} + v_{it})$
where the error term is: $\omega_{it} = \varepsilon_{it} + v_{it}$

Fixed effects: $\mathrm{Y}_{it} = \beta_0 + \beta_1 X_{it} + \mu_i + \mu_t + v_{it}$ where
μ_i is entity specific characteristics omitted by OLS
μ_t is time specific characteristics omitted by OLS
v_{it} is the classical error term

The Hausman statistic is distributed as χ^2 and is computed as:

$$H = (\beta_c - \beta_e)^{'}(V_c - V_e)^{-1}(\beta_e - \beta_e) \tag{20}$$

where

β_e is the coefficient vector from the consistent estimator
β_e is the coefficient vector from the effcient estimator
V_e is the covariance matrix of the efficient estimator
V_e is the covariance matrix of the efficient estimator

The fixed effect regression model takes time and entity specific differences that vary from company to company in consideration. Random effects assume that the entity specific effects are uncorrelated with the other explanatory variables, and the entities are drawn from a population by using random sampling.

3.2.3.3 Analysis of variance (ANOVA)

To explore the relationship between corporate financial performance and ESG we have used one-way ANOVA and Scheffe confidence-interval using the general industry classification standard (GISC) as a group variable. The hypothesis, the test specification and the Scheffe's simultaneous 95% confidence intervals are the following:

$$\mathrm{H}_0 : \mu_1 = \mu_2 = ... = \mu_a \text{ and } \mathrm{H}_1 : \text{ atleast two different} \tag{21}$$

$$\mathrm{F} = \frac{MS_A}{MS_E} \tag{22}$$

$$\mu_1 - \mu_2 = \bar{\mathrm{X}}_1 - \bar{X}_2 \pm \sqrt{(k-1)F_{0,05}} * \sqrt{MS_E} * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{23}$$

where F is the test statistic, $\bar{X}$ is the average values, k is the amount of average values to be compared, n_1 and n_2 is the group populations, $F_{0,05}$ is the degree of freedom for the F-statistic and MS_E is the residual.

3.3 Time series approach using stock portfolios

3.3.1 Portfolio selection criteria

To determine which companies to include in the different portfolios, we have used different proxies for ESG and corporate financial performance. The portfolios are based on companies in the S&P500 and Stoxx 600 indices using monthly data from 2010 to 2018. The risk-free rate used is the 10-year US Treasury yield adjusted to constant maturity.

A common factor used for corporate financial performance is the corporate quality aspect (NBIM, 2015). The quality factor has been defined differently in the literature and different approaches have been used (Novy-Marx, 2013; Fama & French, 2014; Sloan 1996; Piotroski, 2000; Asness, Frazzini & Pedersen, 2014; Graham, 1973). NBIM (2015) defined quality to consist of three categories: profitability, safety and earnings stability/quality. We have used the same approach and constructed quality portfolios by using the following variables: ROA, ROIC, ROE, leverage and the standard deviation of earnings per share. For each variable we have selected the top 33% and the bottom 33% by creating a ranking system. The ranking system sorted the companies from 1 to 3, with 1 being the top 33% and 3 being the bottom 33%, across the different quality variables and summarized the results per company. Companies with the lowest 33% aggregated value were included in the top portfolio, while the companies with the highest 33% aggregated value were included in the bottom portfolio. Companies with high ROA, ROIC, ROE, low leverage and low variation in earnings would be included in the top 33% and vice versa for the bottom 33%. Each portfolio is rebalanced after one year with the same criteria.

Portfolios based on ESG scores are a proxy for sustainable companies and is common in the literature (Landi and Sciarelli, 2019; Verheyden, Eccles and Feiner, 2016; Franzén, 2019). We have also constructed a top and bottom 33% CO2 portfolio based on aggregated direct and indirect CO2 emissions. The purpose of this portfolio is to investigate how CO2 exposure would affect cumulative returns. There are other indicators and variables for environmental exposure, but missing data is a problem and the effective sample size would be small. The complexity in sustainability will be considered in the SEM-models. These portfolios are also rebalanced with the same criteria after one year.

3.3.2 Portfolio risk measures

To analyze the risk-return relationship for the portfolios we have used the portfolio standard deviation of returns, expected return, Sharpe ratio, mean absolute deviation, value at risk and the conditional value at risk (expected shortfall).

The expected portfolio return is calculated by the sum of the security weight multiplied by the expected individual stock return.

$$\mathrm{E}(\mathrm{r}_p) = \sum_{i=1}^{n} w_i E(r_i) \tag{24}$$

The portfolio variance is calculated by the weight of the individual stocks multiplied by the variance-covariance matrix of returns. The portfolio standard deviation is the square root of the portfolio variance.

$$\sigma_p^2 = \sum_{j=1}^{n}\sum_{i=1}^{n} w_j w_i Cov(r_i, r_j) \text{ and } \sigma_p = \sqrt{\sigma^2} \tag{25}$$

The Sharpe Ratio is the return on the portfolio minus the risk-free rate divided by the standard deviation of the portfolio. The ratio is the average return earned in excess of the risk-free rate per unit of volatility or total risk.

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p} \tag{26}$$

The mean absolute deviation (MAD) is the average of the absolute deviation of data points from their mean.

$$\text{MAD} = \frac{1}{N}\sum_{i=1}^{N} | r_i - \mu | \tag{27}$$

The value at risk (VaR) is a risk measure that quantifies the level of risk over a specific period of time. In other words, one measures the potential loss given the probability of occurrence. We have used the Excel-formula PERCENTILE.XCE to calculate VaR using a significance level of 5%. A more general notation to calculate VaR for each portfolio is the following notation:

$$Long:\ p = \mathrm{P}(\Delta V(a) \leq VaR) = F_a(VaR) \text{ and } Short:\ p = 1 - F_a(VaR) \tag{28}$$

Where p= confidence level p , $\Delta V =$ change in asset price for the time period a and $F_a(x)$ is the cumulative distribution of ΔV

Confidence interval VaR
$\mathrm{VaR}_p = x_p = \inf\{x \mid F_a(x) \geq p\}$
inf : The smallest real number

The conditional value at risk is the average of the equal or less observed return for each portfolio given value at risk for the portfolio.

$$\mathrm{ES}_b = \frac{1}{N}\sum_{i=1}^{N} -\min\{r_i - b, 0\} \tag{29}$$

3.4 Granger-causality

Correlation does not equal causality and therefore makes it difficult for econometric models to measure causality directly. The term Granger-causality is used instead of direct causality. A definition for Granger-causality is that a time series x can be said to Granger-cause y if lagged x-values have statistically significant coefficients in the equation for y. This is done to form a better understanding of the environmental, social and governance variables (Lütkepohl, 2005)[1].

To test for Granger-causality one must utilize a VAR model introduced by Sims (1980). This is an extension to the AR model. Using this it is possible to test the relationship between multiple variables at the same time. In this system, each variable has its own equation, and determine whether some variables are exogeneous or endogenous, is done by hypothesis testing. This is essentially testing for Granger-causality.

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{30}$$

Granger-causality from x to y implies that β_{12} is significant different from zero in the equation for y
Granger-causality from y to x implies that β_{21} is significant different from zero in the equation for x

[1] Discusses the use of Granger-causality mainly in chapters 2, 3 and 4.

Granger-causality can also be tested using panel data. Dumitrescu and Hurlin (2012) build on Granger (1969) and developed a test for panel data structure. We have used this test by downloading an add-in, named st0507 (Lopez & Weber, 2017), using STATA 16.0. We have done the same Granger-causality test on different samples due to test limitations. The samples used for testing Granger-causality are based on different business sectors for the environmental, social and governance variables. Due to ESG term complexity we have isolated the environmental, social and governance variables to test for Granger causality. The hypothesis and test statistic for the panel data is denoted as the following:

$$\mathrm{H}_0 : \text{X does not Granger - cause Y and } \mathrm{H}_1 : \text{X Granger - cause Y for one or more panels} \tag{31}$$

$$\tilde{Z} = \sqrt{\frac{N}{2K} * \frac{T-3K}{2K-3}} * \left[\frac{T-3K-3}{T-3K-1} * \bar{W} - K \right] \tag{32}$$

$$\bar{\mathrm{W}} = \frac{1}{N} \sum_{i=1}^{N} W_i \tag{33}$$

where $\tilde{Z}$ is the test statistic, K is the lag order, T is the time variable, N is the sample size and $\bar{\mathrm{W}}$ is the Wald test statistic.

3.4.1 Time period and dataset for testing Granger-causality

The data used for investigating Granger-causality is comprised of data for the companies included in S&P 500 and STOXX 600 indices from 2010 to 2018. This is extracted from Thomson Reuters Datastream and ASSET4. For the description of the variables used in the panel data Granger-causality, see appendix 8.

3.5 Validity and reliability

In this section we will discuss the methodical choices we have done to analyze the relationship between corporate financial performance and ESG. We will discuss the reliability and validity for the Granger-causality test, SEM-models, panel data models and the stock portfolios we constructed. We will also discuss methodical limitations and different types of bias that might affect the results. Validity is about how we can have valid conclusions based on the results, and to what extent the results measures what they are supposed to measure. Reliability is the extent to which results can be reproduced when research is repeated under the same conditions (Ringdal, 2018).

3.5.1 Granger-causality tests

For the Granger-causality panel data tests we must question the validity of the test. The null hypothesis of the test is no Granger-causality, and the alternative hypothesis is that the variable Granger-cause the other variable for at least one company (panel). The test is only based on one dependent continuous variable and one independent continuous variable. Another constrain on this test is the requirement of no missing data and that each panel has variation over the time-period for the given variable. This requirement severely affects the sample size for the environmental and governance factor and the number of variables that is possible to test.

A problem with the full panel data sample, where all business sectors are included, is that the Granger-causality test may become significant due to the large sample size and the test is therefore not very informative. Therefore, we will test on a sector level. However, this can lead to a sample bias. For the selected variables there may also be a selection bias because variables with the least missing data was selected, but no missing data is also a requirement.

3.5.2 SEM-models

For the SEM-models we are facing the same problem compared to Granger-causality test. The full SEM-models requires no missing data to estimate the goodness of fit indices and the path diagram. In terms of validity, we must select variables that have no missing data and be in accordance with existing theory. To ensure validity of the term ESG, we must evaluate if the selected variables measure ESG and not something else.

In terms of reliability for the ESG-CFP SEM-models we have no benchmark model, as far as we know, and based on our observations we find this lacking or non-existent in the literature.

By using cross-sectional SEM-models with some of the same variables we will test different models and see if some variables have a good fit for the different models.

For measuring CFP, we base ourselves on the paper by Hamann et al. (2013), where the dimensions of performance are strictly separated. The alternative would be to utilize some form of hybrid measures, which encompasses two or more dimensions at once. This may have contributed to a less complex model, since you do not end up with so many variables, but the downside is that they are shown to be a less effective measure of performance.

3.5.3 Panel data regression

For the panel data regression models, we account for individual heterogeneity and time specific differences. A fixed or random effects regression might be more valid than pooled ordinary least square regression (OLS) because OLS ignores this individual heterogeneity. In terms of reliability the time-period, 2010-2018, is short, but not all companies have an ESG rating if we were to extend the time-period.

3.5.4 Stock portfolios

The stock portfolios and the risk measures do not contain any information about excess return or abnormal returns but are merely an approach to see the stock market response in relationship to ESG and CFP. The stock market is in a continuous change, and we cannot generalize the results.

4 Analysis

4.1 Introduction

This section will include the analysis of the different tests and the test results. Chapter 4.2 includes descriptive statistics used in the panel data models, SEM-models and panel data regression. Chapter 4.3 includes Granger-causality tests, chapter 4.4 includes CFA-models and chapter 4.5 includes full ESG-CFP SEM-models. Chapter 4.6 includes the panel data regression models and chapter 4.7 includes the stock portfolio approach. In section 4.8 we will discuss the test results and what implications they might have for an investor.

4.2 Descriptive statistics ESG and CFP

This section shows descriptive statistics on an industry (sector) level to get a better understanding of the data the models contain. The purpose of the descriptive approach of the variables on an overall level is to analyze wheter they give any implications for differences between sectors. See appendix 8 for description of the variables.

	ESG Score		Environmental score		Social score		Governance score		Controversies score	
Sector	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Technology	61,17	19,63	63,42	23,99	63,99	20,86	55,37	22,96	40,19	25,32
Consumer Cyclicals	62,4	16,94	64,95	22,3	65,02	19,88	56,5	20,36	42,7	24,25
Financials	62,07	18,05	66,74	23,34	61,55	20,2216	57,51	20,99	40,82	23,98
Healthcare	64,63	16,09	66,31	19,48	67,49	19,54	59,41	21,44	39,24	25,65
Consumer Non-Cyclicals	66,14	14,23	68,18	17,92	69,22	16,56	60,27	20,39	35,18	25,95
Telecommunication Services	61,66	17,42	66,53	21,64	61,26	20,4	56,69	19,36	37,5	27,77
Energy	65,47	16,31	67,21	21,29	67,95	18,95	60,63	20,49	37,39	25,34
Industrials	61,28	17,2	63,52	21,88	63,37	19,41	56,36	22,51	41,12	24
Utilities	63,56	14,77	65,82	19,26	64,88	17,01	59,52	19,98	37,12	25,05
Basic Materials	63,15	18,44	64,07	21,97	65,45	22,22	59,46	21,85	39,78	24,67
Average score across all sectors	62,78	17,28	65,675		65,018		58,172		39,104	

Table 4.1 Descriptive statistics ESG, environmental, social, governance and controversies score

Table 4.1 shows the ESG rating on a sector level for the S&P 500 and Stoxx 600 companies. The ESG score is an aggregated score based on all ESG data in the Thomson Reuters Eikon database with equal weights for the environmental, social and governance data. All the scores range from 0-100, and the closer to 100, the better. The table shows that average ESG, environmental, social and governance score, and the standard deviation for all sectors is not that different. However, using ANOVA and Scheffe 95% confidence interval shows that there are differences across sectors. See table 4.2.

The ESG score is sensitive to the controversies score. The controversies score is defined by Thomson Reuters Eikon as negative media coverage for the companies based on global media. The score is calculated by using controversies measures with a weighted average. If a scandal happens the company will be penalized, and development in the scandal will affect the score in later periods e.g. if a lawsuit happens (Refinitive, 2020). Using ANOVA and Scheffe 95% confidence interval indicates that the controversies score is significantly different between consumer cyclicals and consumer non-cyclical and industrials and consumer non-cyclicals. However, we must be critical to how controversies score is measured and how it captures controversies.

Sector	Technology	Consumer Cyclicals	Financials	Healthcare	Consumer Non-Cyclicals	Telecommunication Services	Energy	Industrials	Utilities	Basic Materials
Technology	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C
Consumer Cyclicals	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C
Financials	ESG E S G C	ESG E S* G* C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C
Healthcare	ESG E S* G C	ESG E S* G* C	ESG E S* G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C
Consumer Non-Cyclicals	ESG* E S G C	ESG* E* S* G* C*	ESG* E S* G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E* S G C	ESG E S G C	ESG E S G C
Telecommunication Services	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S* G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C
Energy	ESG* E* S* G*	ESG* E* S G* C	ESG* E S* G C	ESG E S G C	ESG E S G C	ESG E S* G C	ESG E S G C	ESG E S G C	ESG E S G C	ESG E S G C
Industrials	ESG E S G C	ESG E S G C	ESG E* S G C	ESG* E S* G C	ESG* E S* G* C*	ESG E S G C	ESG* E* S* G* C	ESG E S G C	ESG E S G C	ESG E S G C
Utilities	ESG* E S G* C	ESG* E S G* C	ESG* E S G* C	ESG E S G* C	ESG E S* G C	ESG E S G* C	ESG E S G C	ESG* E S G* C	ESG E S G C	ESG E S G C
Basic Materials	ESG E S G* C	ESG E S G* C	ESG E S G C	ESG E S G C	ESG E S* G C	ESG E S G C	ESG E S G C	ESG* E S G* C	ESG E S G C	ESG E S G C

*)Significant on a 5% level

Table 4.2 Matrix of results based on Scheffe's 95% confidence interval for ESG, environmental, social, governance and controversies score using economic sector as a group variable.

Looking at the individual environmental, social and governance score we also find sector differences by using ANOVA and Scheffe's 95% confidence interval. Data indicates that the environmental score is significantly different between energy and technology, consumer cyclicals and consumer non-cyclicals, financials and industrials, industrials and energy and industrials and consumer non-cyclicals. For the social factor, table 4.2 indicates that the social score is significantly different for 14 combinations between all the sectors. Regarding the governance score, table 4.2 shows that the score is significantly different for 15 combinations between all the sectors. In other words, there is an indication that there are significant differences between sectors for the social and governance score.

	Return on assets		Return on invested capital		Return on equity		Stock return	
Sector	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Technology	11,17	22,95	24,25	127,476	34,17	144,48	19	35,58
Consumer Cyclicals	8,49	7,62	13,7	13,83	22,49	63,37	15,65	35,15
Financials	5,21	6,99	8,82	26,87	16,4	65,44	10,07	27,98
Healthcare	7,87	9,06	11,52	15,75	17,98	27,78	18,71	35,83
Consumer Non-Cyclicals	7,48	8,17	11,33	13,67	24,93	77,64	11,98	26,18
Telecommunication Services	6,19	5,37	8,81	7,35	26,8	134,63	4,56	24,17
Energy	5,87	8,32	9,11	13,08	15,07	31,41	10,36	40,09
Industrials	6,94	5,73	11,51	17,48	22,8038	77,88	13,08	29,48
Utilities	5,41	5,9	8,15	14,43	13,83	13,83	9,12	26,76
Basic Materials	7,31	8,57	10,56	12,6	14,12	19,45	11,08	33,24
Average all sectors	7,22	10,1	12,03	43,03	20,76	74,13	13,16	31,98

Table 4.3 Descriptive statistics indicators for CFP

Table 4.3 shows some indicators of corporate financial performance based on Hamann et al. (2013) and NBIM (2015). The indicators, ROA, ROIC and ROE, measure profitability for the selected sectors. Data indicate that the technology sector has the highest average values for ROA, ROIC, ROE and stock return. The standard deviations for the technology sector are high compared to other sectors and indicate variation within the sector.

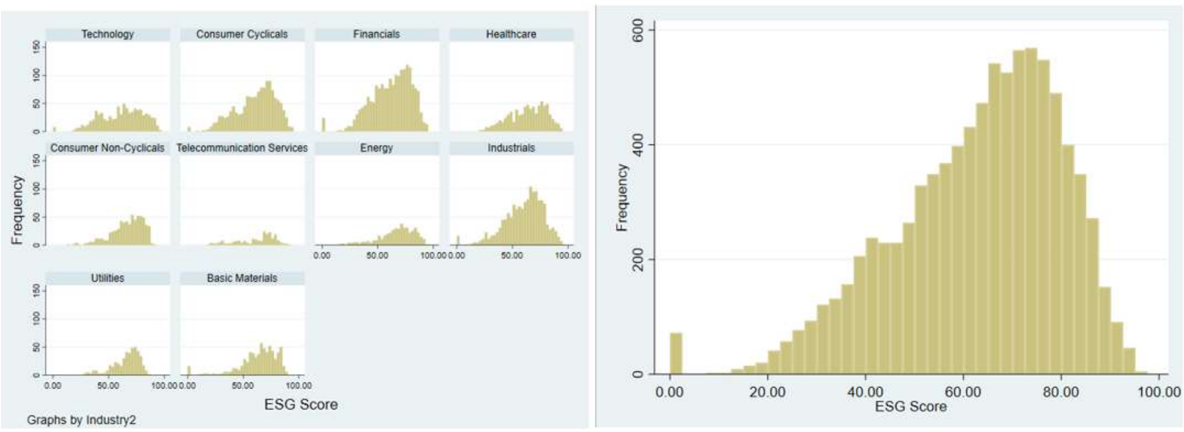


Figure 4.1 Histogram: Distribution ESG score by sector and all sectors

Figure 4.1 shows a frequency histogram for the ESG score on a sector and overall level. The figure shows that all the sectors have a left skewed distribution except for technology, telecommunication services and basic materials. Appendix 11 shows the environmental, social, governance, and controversies score. The diagram for controversies score shows that all the scores are either very high or low and nothing in between.

4.3 Granger causality

The Granger causality test utilizes the data structure which was presented earlier. In part 4.3.1 the results from the tests will be analyzed, and in section 4.3.2 a discussion of the implications will follow which will help us form a better understanding of the variables. Churet and Eccles (2014) found that there is a difference between sectors in relationship to what variables are reported. Sassen, Hinze and Hardeck (2016) find that the aggregated scores in the social dimensions hide a more nuanced picture of the connections between ESG and CFP, and by disaggregating the scores they find more information on what these connections are. Building on their findings, the individual variables are examined, and the business sectors are separated. The main hypothesis for this test is that different combinations of variables will be significant for some of the business sectors. In other words, it is not expected that a combination of variables is to be significant for all sectors. The social variables were tested with 835 companies, while the environmental variables were tested with 186 companies. None of the governance variables could be tested because of a lack of variation over time for the variables. Stationary variables are a requirement for the Granger causality tests, and all variables seem to be stationary, see appendix 3 for test results.

4.3.1 Results Granger causality tests

To get some indication on how some of the ESG data are possibly connected, a Granger causality test is performed on the panel data. One of the problems of checking panel data for Granger causality is that there are not many tests that consider the panel structure of the data. Since it is of interest to see the effect between variables and keeping the panel structure, a test for Granger causality in panel data developed by Lopez and Weber (2017) was used. This test is based on a paper explaining Granger causality in panel data written by Dumitrescu and Hurlin (2012), and is explained in chapter 3.4. The test checks if there are one or more panels that have a significant Granger causality between the independent and dependent variable, and because of this you end up with many significant results if you include many panels in the dataset. For example, if all the panels are combined and tested for Granger causality, the results will be significant on the 5% level for all combinations of variables shown here in table 4.4 and 4.5. To combat this somewhat, the companies are separated into sectors and the test is run on each sector individually. The variables have been tested both ways, and the results are given in table 4.4 to table 4.8.

		LNWaterWithdrawal on LNco2Direct		LNco2Direct on LNWaterWithdrawal		LNco2Direct on LNEnergyUse		LNEnergyUse on LNco2Direct	
Sector	Panels	Z tilde	P-value	Z tilde	P-value	Z tilde	P-value	Z tilde	P-value
Technology	9	0,5100	0,6100	0,2189	0,8268	-0,0971	0,9227	1,8284	0,0675*
Consumer Cyclicals	26	0,7804	0,4352	-0,1474	0,8828	1,1783	0,2387	1,5829	0,1134
Financials	37	1,5313	0,1257	0,8784	0,3797	2,0841	0,0371**	3,0577	0,0022**
Healthcare	23	3,4467	0,0006**	0,3007	0,7637	1,1368	0,2556	7,9213	0,0000**
Consumer Non-Cyclicals	21	3,3831	0,0007**	0,5077	0,6117	1,2614	0,2071	-0,1415	0,8874
Telecommunication Services	9	-0,3958	0,6922	-0,7343	0,4628	3,6385	0,0003**	-0,1697	0,8653
Energy	11	1,2184	0,2231	1,4770	0,1397	0,5510	0,5816	4,5958	0,0000**
Industrials	26	4,7086	0,0000**	7,4338	0,0000**	1,3781	0,1682	0,2224	0,8240
Utilities	8	0,9206	0,3573	-0,2283	0,8194	0,0924	0,9263	-0,2164	0,8287
Basic Materials	16	1,1807	0,2377	1,0603	0,2890	-0,1412	0,8877	0,4139	0,6790

**) Singnificant on a 5% level
*) Significant on a 10% level

Table 4.4 Results Granger-causality tests environmental variables

The results in table 4.4 show the Z-value and p-value regarding Granger causality for LNco2direct emissions per sector and LNWaterWithdrawal, and additionally LNEnergyUse with LNco2Direct. The result here indicates that the connection between LNco2Direct and LNEnergyUse is stronger for a higher number of the sectors than LNco2Direct and LNWaterWithdrawl, because there are more significant results for the former than the latter.

Sector	Panels	LNEnergyUse on LNWasteTotal		LNWasteTotal on LNEnergyUse		LNco2Direct on LNWasteTotal		LNWasteTotal on LNco2Direct	
		Z tilde	P-value	Z tilde	P-value	Z tilde	P-value	Z tilde	P-value
Technology	9	0,6675	0,5044	-0,1377	0,8905	1,3451	0,1786	0,6876	0,4917
Consumer Cyclicals	26	-0,5957	0,5514	-1,1080	0,2679	-0,2027	0,8394	0,6723	0,5014
Financials	37	8,7535	0,0000**	-0,0919	0,9268	12,2229	0,0000**	0,4408	0,6593
Healthcare	23	5,0425	0,0000**	0,5444	0,5862	1,8721	0,0612*	2,0501	0,0404**
Consumer Non-Cyclicals	21	1,6188	0,1055	5,6943	0,0000**	0,6132	0,5398	6,0826	0,0000**
Telecommunication Services	9	0,3740	0,7084	-0,2949	0,7680	-0,5348	0,5928	1,0015	0,3166
Energy	11	0,3601	0,7188	2,9403	0,0033	1,3966	0,1625	0,8658	0,3866
Industrials	26	0,6993	0,4844	2,4336	0,0149**	6,6001	0,0000**	1,1988	0,2337
Utilities	8	-0,5747	0,5655	-0,0650	0,9481	-0,3859	0,6996	-0,6742	0,5002
Basic Materials	16	1,0750	0,2824	7,1577	0,0000**	0,4286	0,6682	-0,3129	0,7544

**) Singnificant on a 5% level
*) Significant on a 10% level

Table 4.5 Results Granger-causality tests environmental variables

In table 4.5 we see the tests done on LNWasteTotal with LNEnergyUse and the combination LNWasteTotal with LNco2Direct. The differences between sectors seem to be present also for these combinations of environmental variables. The results are very similar, where the sectors that are significant for energy use and waste also are significant for LNco2Direct and LNWasteTotal, except for the sector basic material, which has a significant result only for LNWasteTotal on LNEnergyUse.

Sector	Panels	LNEnergyUse on LNWaterWithdrawal		LNWaterwithdrawal on LNEnergyUse		LNWasteTotal on LNWaterWithdrawal		LNWaterWithdrawal on LNWasteTotal	
		Z tilde	P-value	Z tilde	P-value	Z tilde	P-value	Z tilde	P-value
Technology	9	-0,1473	0,8829	-0,5522	0,5808	0,8254	0,4091	-0,3414	0,7328
Consumer Cyclicals	26	1,2717	0,2035	0,2321	0,8165	0,1145	0,9088	-0,3817	0,7027
Financials	37	2,8412	0,0045**	1,3800	0,1676	2,1133	0,0346**	9,7503	0,0000**
Healthcare	23	0,6593	0,5097	5,6585	0,0000**	0,9671	0,3335	1,3460	0,1783
Consumer Non-Cyclicals	21	4,4533	0,0000**	0,4431	0,6577	-0,1436	0,8858	0,6549	0,5125
Telecommunication Services	9	-0,1835	0,8544	3,0130	0,0026**	3,2429	0,0012**	-0,2421	0,8087
Energy	11	-0,4130	0,6796	0,1014	0,9192	0,1587	0,8739	-0,1886	0,8504
Industrials	26	-0,4303	0,6669	4,4595	0,0000**	0,0010	0,9992	1,4373	0,1506
Utilities	8	-0,1414	0,8876	5,6613	0,0000**	-0,3341	0,7383	-0,3665	0,7140
Basic Materials	16	1,2060	0,2278	-0,0158	0,9874	0,0343	0,9726	0,3695	0,7117

**) Singnificant on a 5% level
*) Significant on a 10% level

Table 4.6 Results Granger-causality tests environmental variables

Table 4.6 highlights the Granger causality test results for the variables concerning LNEnergyUse and LNWaterWithdrawal together with LNWasteTotal and LNWaterWithdrawal. The effect between LNEnergyUse and LNWaterWithdrawal shows some significant results, and exhibits the highest number of significant effects for the tested environmental variables, with six different sectors showing significant results on the 5% level. For LNWaterWithdrawal and LNWasteTotal, there are not many significant results with only two sectors producing significant results on the 5% level.

To have a more nuanced look at the variables concerning governance and the social dimension the same Granger causality test can be applied. The test is strict in the form that it requires variation over the time-period and no missing values. Since many of the variables under the governance and social dimension are dichotomous variables, these were impossible to perform the test on. For the governance dimension the prerequisites of the test effectively eliminated all pure dichotomous governance variables. Instead the test is used on score variables from Thomson Reuters Asset4. Table 4.7 contains the results for the social variables that could be tested, while table 4.8 contains the results for the score variables.

		ProductResp. on HumanRightsScore		HumanRightsScore on ProductResp.	
Sector	Panels	Z tilde	P-value	Z tilde	P-value
Technology	72	4,9987	0,0000**	4,5272	0,0000**
Consumer Cyclicals	130	3,8126	0,0001**	0,8116	0,4170
Financials	185	15,6242	0,0000**	21,0316	0,0000**
Healthcare	71	0,7554	0,4500	1,5556	0,1198
Consumer Non-Cyclicals	67	-0,6902	0,4901	0,8975	0,3695
Telecommunication Services	21	1,5699	0,1164	-0,6079	0,5421
Energy	47	2,1756	0,0296**	0,7261	0,4678
Industrials	137	26,8447	0,0000**	1,4201	0,1556
Utilities	41	3,3057	0,0009**	1,7481	0,0805*
Basic Materials	64	-0,3352	0,7375	6,6085	0,0000**

**) Singnificant on a 5% level
*) Significant on a 10% level

Table 4.7 Results Granger-causality test social variables

Sector	Panels	ESGControversies on CSRStrategy		CSRStrategy on ESGControversies		CSRStrategy on ESGScore		ESGScore on CSRStrategy	
		Z tilde	P-value	Z tilde	P-value	Z tilde	P-value	Z tilde	P-value
Technology	74	3,4160	0,0006**	7,5280	0,0000**	1,7925	0,0731*	2,4976	0,0125**
Consumer Cyclicals	130	-0,1505	0,8804	0,9524	0,3409	1,0050	0,3149	8,1594	0,0000**
Financials	186	2,0039	0,0451**	0,3618	0,7175	3,1118	0,0019**	6,1971	0,0000**
Healthcare	71	5,8163	0,0000**	-0,1557	0,8763	1,9429	0,0520*	5,6233	0,0000**
Consumer Non-Cyclicals	67	-0,3785	0,7050	-0,2231	0,8235	0,7007	0,4835	0,9255	0,3547
Telecommunication Services	21	0,0728	0,9419	-1,0001	0,3173	0,9637	0,3352	1,4472	0,1478
Energy	47	-0,2086	0,8348	-0,5546	0,5792	0,2247	0,8222	2,4186	0,0156**
Industrials	138	3,3916	0,0007**	1,2719	0,2034	0,8828	0,3774	3,1321	0,0017**
Utilities	41	-0,2921	0,7702	0,7973	0,4253	0,0692	0,9448	-0,3500	0,7264
Basic Materials	61	0,8424	0,3996	-0,1542	0,8775	2,5660	0,0103**	4,0173	0,0001**

**) Singnificant on a 5% level
*) Significant on a 10% level

Table 4.8 Results Granger-causality tests score variables

Table 4.7 shows that six out of ten sectors have a significant result for product responsibility score on human rights score, and four out of ten sectors have significant results for human rights score on product responsibility. The relationship between the product responsibility score and the human rights score seems to be existent if one looks at both possible combination of the variables. This is in some ways contradictory to our hypothesis since the results for most sectors are significant one way or another. Only three sectors do not have any significant results here, namely consumer non-cyclicals, telecommunication services and energy.

Table 4.8 highlights some of the score variables available from Thomson Reuters Asset4. ESG controversies score and CSR strategy score exhibit some significant results for different business sectors. There are more significant results regarding the ESG score combined with the CSR strategy score. This also seems to be significant for many of the sectors and a bit contradictory to our hypothesis. No certain conclusion can be drawn by using this test if there are one-way, two-way or another form of interaction effect between the variables.

4.3.2 Implications of Granger causality results

The results indicate some connection between the different variables because some of them has a Granger causality effect in certain sectors. There is evidence for this across both the environmental and social dimension, in addition to the scoring variables.

For the environmental dimension there is a relatively high number of significant sectors for Granger causality between CO2 emissions and energy use. From a logical standpoint this can be explained by the fact that companies that have a larger energy use also have a larger emission of CO2. A higher or lower value of either lagged variable is accompanied with a higher or lower value for the other in the next period. This may indicate that there is some connection between these variables, and that this should be taken into consideration when modelling for the relationship between ESG and CFP. This can also have an effect for the SEM-model since prior research has found environmental variables to influence the financial performance. Hassan & Romilly (2018) found a negative relation between what they called greenhouse gas emissions and economic performance, implying that higher emissions lead to lower economic performance.

Energy use and water withdrawal also have some significant results for different sectors, indicating that water withdrawal or energy use in the previous period affects the other in the period after. It does not seem to be as many significant results for the other variables in the environmental dimension. There seems to be a difference between sectors here as well, but this difference cannot be isolated in this analysis alone. The reason for this is that there is a relatively large difference in the number of companies for each sector, and that this can be some of the cause of the difference. The test specification may have a low test power with many panels.

In the social dimension, the variables product responsibility score and human rights score show many significant results for the different sectors. The interpretation of this is that there is a connection between the lagged value of one of the variables and the other. This seems to indicate that the extent to which a company produces responsible products is linked to how they follow human rights for some sectors. The ESG score seems to have a one directional effect on CSR strategy, except for the sectors technology, financials, healthcare, and basic material where the effect seems to be bidirectional.

Across all the different variables that have been tested here, the Granger causality connection differs between the business sectors. This shows that there may be some differences in how different ESG variables affect the business sectors, and that one should consider the sector up against the ESG variables when looking at ESG investing.

4.4 SEM-models

To form a broader picture of what the relationship the different ESG factors have towards CFP, SEM-models have been constructed. To analyze the relationship, an explorative and confirmatory factor analysis of selected ESG variables is conducted. This will serve as a starting point for the structural equation models.

The purpose of the factor models is to analyze the relationship among the measured variables to determine whether the observed variables can be grouped into a smaller set of underlying factors or theoretical constructs (Thompson, 2004; Worthington & Whittaker, 2006). The confirmatory factor analysis is built on the results of the explorative factor analysis. However, the explorative factor analysis does not show a clear pattern of variables that should be included in a factor for the governance and social factors, as shown in appendix 7. For the environmental factor, the explorative factor analysis shows a clear pattern for the variables related to energy/resource use. The explorative and confirmatory factor analysis is based on data from 2018. The path diagrams for the full SEM-models can be seen in appendix 5. The full overview and description of the variables can be found in appendix 8. The two-sided critical values for the t-tests, 10%, 5% and 1%, are 1,658, 1,980 and 2,617 (Studenmund, 2017). For the largest sample, N=421, the critical values are 1,645, 1,96 and 2,576.

4.4.1 Confirmatory factor analysis environmental, social and governance

Measurement model observed X-variables

	Indicator	Ksi	Standardized factor loading	t-value	R2
Environmental	LN WaterWithdrawl	Ksi1	0,81	17,368	0,65
N=268	LNCO2 emissions direct		0,83	15,56	0,69
	LN Non Hazardous waste		0,83	13,319	0,69
	LN Hazardous waste		0,73	10,638	0,54
	Eco design products	Ksi2	0,2	1,72	0,04
	Animal testing		0,34	1,969	0,11
	Renewable/clean energy products		-0,5	-1,859	0,25
Social	Health & safety policy	Ksi1	0,6	2,746	0,36
N=319	Policy diveristy and opportunity		0,2	2,072	0,04
	Strikes		0,09	2,193	0,01
	Policy Human rights		0,57	3,969	0,32
	Policy fair competition	Ksi2	0,74	4,9	0,55
	Policy bribery and corruption		0,63	3,461	0,4
	Whistleblower protection		0,35	3,413	0,13
	Policy Customer health & safety	Ksi3	0,7	4,44	0,48
	Policy data privacy		0,04	0,571	0,002
	Policy responsible marketing		0,24	2,98	0,06
	Policy fair trade		0,15	1,918	0,02
Governance	Audit commitee independence	Ksi1	0,8	9,375	0,63
N=210	Audit commitee non executive		0,17	1,573	0,03
	Compensation commitee independence		0,89	9,482	0,79
	Compensation commitee non executive		0,65	5,361	0,42
	Board structure gender diversity	Ksi2	0,29	4,64	0,15
	Board specific skills		0,52	-7,369	0,27
	Board members strictly independent		-0,6	-8	0,36
	CSR sutainability commitee	Ksi3	0,46	2,241	0,23
	CSR reporting global activites		0,33	1,891	0,11
	CSR sustainability external audit		0,52	2,534	0,27
	ESG reporting scope		0,27	1,598	0,07

Table 4.9 CFA measurement model for observed indicators for environmental, social and governance

Table 4.9 shows the standardized factor loadings, t-values and the R^2 for the confirmatory factor analysis. The factor loading is standardized between –1 and 1 and is an estimate of the path coefficient depicting the effect of a factor on an item or manifest variable (Bowen & Guo, 2012). The table shows that the environmental factor, Ksi1, has a high standardized factor loading, and most of the variables are significant with t-values > $|1,98|$. The R^2 is an indicator of reliability and shows how much of the variance in the observed indicators can be explained by the latent factors. R^2 above 0,5 is considered high, and values between 0,35 and 0,5 are considered being moderate. The R^2 for the equations are also high and may indicate good reliability. The second environmental factor, Ksi2, has very low standardized factor loadings and R^2 for the equations, and only one variable is significant on a 5% level.

Table 4.9 also shows that the results are mixed for the social factor with some variables having a high standardized factor loading, and some are very low (close to zero). Most of the variables are significant with a t-value > $|1,98|$, but the R^2 for the equations is very low and may indicate low reliability. For the governance factor the table shows that the indicators for Ksi1 mostly have high standardized factor loadings and significant variables. The R^2 for the equations is mostly high for Ksi1 and low for Ksi2 and Ksi3.

4.4.2 CFA ESG measures of reliability

		Measures of reliability		
	Latent factor	Composite reliability	Average Variance extracted	Cronbach's Alpha
Environmental	Ksi1= Rescource use	0,8767	0,6406	0,8880
	Ksi2=Environmental products	0,0006	0,1349	-0,0782
Social	Ksi1= Labor rights	0,3946	0,1848	0,3660
	Ksi2= Workforce protection	0,6064	0,3572	0,5869
	Ksi3= Corporate social policies	0,2707	0,1422	0,2526
Governance	Ksi1= Independence	0,7473	0,4693	0,6637
	Ksi2= Board composition	0,0413	0,2598	-0,0419
	Ksi3= Sustainable reporting	0,4354	0,1705	0,5279

Table 4.10 Measures of reliability CFA environmental, social and governance factors

To measure the reliability of the latent factors we have used three measures of reliability, see appendix 6 for notation. Cronbach's Alpha (CA) is a measure of internal reliability, and a value of 0,7 is considered as a lower boundary for acceptable reliability (Ringdal, 2018). The composite reliability (CR) measures term reliability and is an indicator of the shared variance among observed variables used as an indicator of a latent construct (Fornell & Larcker, 1981). The average variance extracted (AVE) also uses the measurement errors and standardized factor loadings and is also an indicator of term reliability (Bagozzi & Yi, 1988). A rule of thumb for the reliability measures is that CR > 0,6 and AVE > 0,5 indicates good reliability, but no absolute cut-off values.

Table 4.10 shows that CR, CA and AVE indicates good reliability with values higher than the rule of thumb for Ks1 for the environmental CFA. Ksi2 for the environmental CFA indicates low reliability with very low values for AVE and CR. The negative CA is due to negative average correlation and indicate low internal reliability. The negative CA is also an indication that the selected variables are not fit indicators for the latent construct. Ksi1 and Ksi3 for the social factor have values below the rule of thumb and does not show acceptable reliability. CR for Ksi2 for the social factor indicates acceptable fit, but the rest of the reliability measures do not. For the governance factor, CR shows acceptable fit and CA and AVE do not deviate much from having an acceptable fit for Ksi1. Ksi2 and Ksi3 for governance do not show acceptable fit.

4.4.3 CFA environmental, social and governance goodness of fit indices

Goodness of fit indices	Environmental	P-value	Social	P-value	Governance	P-value
C1	30,311	0,0014	59,755	0,0293 *	42,681	0,3566 **
C3	35,179	0,0002	58,719	0,0358 *	39,374	0,4982 **
RMSEA	0,0809	0,0632 *	0,0379	0,83 **	0,0179	0,936 **
NFI	0,952 **		0,758		0,943 *	
NNFI	0,935		0,873		1,0001 **	
PNFI	0,499		0,565		0,686	
CFI	0,966 *		0,905		1 **	
IFI	0,967		0,912		1	
RFI	0,909		0,675		0,922	
Critical N	188,663		352,748		339,075	
RMR	0,0922		0,00891		11,372	
SRMR	0,0707 *		0,0483 **		0,0471 **	
GFI	0,97 *		0,967 **		0,965 **	
AGFI	0,924 **		0,946 **		0,942 **	
PGFI	0,381		0,6		0,585	
Test for multivariate normality	Test statistic	P-value	Test statistic	P-value	Test statistic	P-value
Skewness	Z=20,777	0,000	Z=82,890	0,000	Z=970,509	0,000
Kurtosis	Z=6,045	0,000	Z=22,472	0,000	Z=27,865	0,000
Skewness & kurtosis	Chi2=468,234	0,000	Chi2=7375,744	0,000	Chi2=18350,739	0,000

**) Good fit
*) Acceptable fit

Table 4.11 CFA goodness of fit indices environmental, social and governance

To determine goodness of fit for the CFA-models we have used fit indices we described in chapter 3 based on Schermelleh-Engel et al. (2003), see table 4.11. All the models reject the null hypothesis of multivariate normality. To handle the violation of the assumption of multivariate normality the models are estimated using Robust Maximum Likelihood (RML). RML is also robust against a not-correctly specified model (Schermelleh-Engel et al., 2003). The CFA models for environmental and social factors reject the null hypothesis of exact fit for the χ^2 test. Due to the non-normality C3 is the more reliable χ^2, but the tables show that the value of C3 is very close to C1. RMSEA and the test for close fit indicate acceptable fit for the environmental CFA and good fit for social and governance CFA. NFI, NNFI, CFI, IFI and RFI close to a value of 1 indicate good fit and PGFI and PNFI are used for model comparison. GFI, AGFI and SRMR show acceptable and good fit for the models. The incremental indices, NFI, NNFI and CFI, show good and acceptable fit for governance and the environmental CFA. For the social CFA, the model indicates some disturbance in the data due to the difference between GFI and RFI. The values of the incremental indices are below GFI and indicates noise. The environmental model has a Critical N < 200 and indicates bad fit (Hoelter, 1983).

4.4.4 CFA corporate financial performance

The CFA for CFP is based on Hamann et al. (2013). They conducted a CFA for CFP and constructed 4 latent factors with good statistical properties. They operationalized CFP as liquidity, profitability, growth, and stock market performance. Hamann et al. (2013) used a very large sample size for their different models. Our approach to CFP is using the same latent constructs for profitability and growth, but different indicators for profitability. ROE and ROIC are used instead due to focus on the quality factor. Stock market performance with different indicators compared to Hamman et al. (2013) is also tested. An explorative factor analysis is also conducted, see appendix 7, which the CFA is based on.

4.4.5 CFA corporate financial performance reliability

CFA CFP Measurement model observed X-variables

Model	Indicator	Ksi	Standardized factor loading	t-value	R2
CFP1	ROA	Ksi1	0,92	3,529	0,843
N=1022	ROE		0,42	3,461	0,176
	ROIC		Fixed 1	2,505	Fixed 1
	Net 1-year sales/revenue growth	Ksi2	0,66	3,703	0,434
	1-year total asset growth		0,78	5,28	0,606
	1-year employee growth		0,88	3,402	0,778
CFP2	ROA	Ksi1	0,37	1,698	0,136
N=1022	ROE		Fixed 1	4,077	Fixed 1
	Net 1-year sales/revenue growth	Ksi2	0,66	3,703	0,434
	1-year total asset growth		0,78	5,28	0,606
	1-year employee growth		0,88	3,402	0,778
CFP3	ROA	Ksi1	0,96	3,676	0,926
N=151	ROE		0,76	2,897	0,573
	ROIC		0,98	3,984	0,956
	Net 1-year sales/revenue growth	Ksi2	0,39	3,363	0,149
	1-year total asset growth		0,76	4,192	0,571
	1-year employee growth		0,77	4,466	0,592
CFP4	ROA	Ksi1	0,96	3,579	0,927
N=140	ROE		0,75	2,756	0,561
	ROIC		0,98	3,907	0,956
	Net 1-year sales/revenue growth	Ksi2	0,4	3,102	0,164
	1-year total asset growth		0,79	4,058	0,619
	1-year employee growth		0,74	4,168	0,541
	Annual Stock return	Ksi3	0,87	4,384	0,748
	Price/earnings ratio		0,52	1,216	0,27
	Dividend yield		-0,31	-5,694	0,094

Table 4.12 CFA CFP measurement model for observed indicators

Table 4.12 shows the standardized factor loadings, t-values and R^2 for the CFA of CFP. Due to having little missing CFP data, the CFP indicators have been tested using a full sample and the sample used on the full SEM-model for ESG-CFP. The purpose of this is to increase the test power. The fixation in model CFP1 is due to a negative measurement error very close to zero, and CFP2 is due to not being able to estimate the measurement error, Θ_δ, for ROE. The negative measurement error is usually an indication of disturbance in the data and/or a poor fit for the indicator. However, the table shows that Ksi1 does not have the same problem in the smaller samples. This raises the question why this results in a slightly negative measurement error in the large sample, but not in the small sample. This could be a result of the selected companies in the small sample or that the indicators are not robust for the largest sample. Either way, the tables show that the standardized factor loadings are high/moderate and significant on a 5% level with t-values > |1,98| and |1,96|. R^2 seems to be affected by the fixation in model CFP1 and CFP2.

Table 4.12 shows that the selected indicators for Ksi2 have moderate/high standardized factor loadings (absolute value) for all the models. The indicators are significant on a 5% level and the R^2 ranges from low to high. R^2 for Net 1-year sales/revenue growth is low in the small sample and moderate in the large sample. Model CFP4 includes proxies for stock market performance (Ksi3). The table shows that the price/earnings ratio is not significant on a 5% level while stock return and dividend yield are. R^2 is high for annual stock return and low for the P/E-ratio and dividend yield.

4.4.6 CFA corporate financial performance goodness of fit indices

Goodness of fit indices	CFP1	P-value	CFP2	P-value	CFP3	P-value	CFP4	P-value
C1	19,522	0,0341*	6,502	0,3693**	8,954	0,4415**	50,695	0,0038
C3	26,232	0,0034	20,433	0,0023	14,298	0,1121**	45,762	0,0135*
RMSEA	0,0306**	0,945**	0,00905**	0,985**	0,00**	0,7**	0,0792*	0,0786*
NFI	0,992**		0,984**		0,974**		0,922*	
NNFI	0,992**		0,981**		0,983**		0,954*	
PNFI	0,661		0,59		0,584		0,691	
CFI	0,995**		0,59		0,99**		0,966*	
IFI	0,995		0,989**		0,99		0,966	
RFI	0,988		0,973		0,956		0,895	
Critical N	904,37		841,062		228,302		143,647	
RMR	21,77		24,848		18,11		198,996	
SRMR	0,0133**		0,0141**		0,0636*		0,102	
GFI	0,994**		0,997**		0,981**		0,923*	
AGFI	0,987**		0,994**		0,956**		0,872*	
PGFI	0,473		0,339		0,42		0,554	
Test for multivariate normality	Test statistic	P-value	Test statistic	P-value	Test statistic	P-value	Test statistic	P-value
Skewness	Z=247,340	0,000	Z=189,998	0,000	Z=39,612	0,000	Z=61,112	0,000
Kurtosis	Z=1895,40	0,000	Z=49,281	0,000	Z=15,472	0,000	Z=17,394	0,000
Skewness & kurtosis	Chi2=64120,994	0,000	Chi2=38527,990	0,000	Chi2=1808,532	0,000	Chi2=4037,214	0,000

**) Good fit
*) Acceptable fit

Table 4.13 CFA goodness of fit indices for CFP

Table 4.13 shows that all the models reject the null hypothesis of multivariate normal distribution, and based on this the models are estimated using RML. The table shows that most of the indices show good or acceptable fit. C3 is the more reliable χ^2 in a small sample, but there is only a small difference between C1 and C3 indicating that the non-multivariate normality might not be a problem. The indices do not show a large gap between NFI and GFI and RFI, except for CFP4, and indicates that there is little disturbance in the data for CFP1-CFP3. A critical N < 200 indicates poor fit for model CFP4 and for the large samples CN >

400 is the cutoff value indicating good fit for CFP1 and CFP2. Overall, the indices show good fit for CFP1-CFP3.

Measures of reliability					
Model	Latent factor	Ksi	Composite reliability	Average Variance extracted	Cronbach's Alpha
CFP1	Profitability	Ksi1	0,85	0,67	0,80
	Growth	Ksi2	0,82	0,61	0,82
CFP2	Profitability	Ksi1	0,69	0,57	0,54
	Growth	Ksi2	0,82	0,48	0,82
CFP3	Profitability	Ksi1	0,93	0,82	0,90
	Growth	Ksi2	0,69	0,44	0,66
CFP4	Profitability	Ksi1	0,93	0,82	0,90
	Growth	Ksi2	0,69	0,44	0,66
	Stock market performance	Ksi3	0,38	0,37	0,07

Table 4.14 CFA measures of reliability for CFP

Table 4.14 shows the reliability measures for the latent constructs. The measures of reliability show that all the hypothesized latent constructs except stock market performance has a CR above 0,6. The tables also show that most of the latent constructs has an AVE higher or close to 0,5. Cronbach's alpha is mostly higher or close to 0,7 for the latent constructs except for stock market performance. The measures of reliability for stock market performance does not indicate good reliability. Overall, the CFA shows good fit and reliability for growth and profitability, but sample size might affect the results. Model CFP4 is favored by PNFI and PGFI but based on the low reliability the stock market performance indicator it will be dropped in the full ESG-CFP SEM-models.

4.5 Full SEM ESG-CFP and E-CFP models

4.5.1 Description full SEM-models

In this section full SEM-models have been constructed for the relationship between corporate financial performance and environmental, social and governance factors using an explorative approach based on MSCI (2020a) and the CFA. The models are based on data from 2018. The models have not been tested for other years because the focus has been on trying to create a reliable measurement instrument before testing it on other time periods. A lot of different models have been tested and the more complex models severely constrains the sample size and limits what is possible to test. The latent constructs profitability and growth are used as proxies for corporate financial performance. For the indicators/proxies for the latent constructs E, S and G many variables have been used, but many dichotomous variables have been omitted due to having only one value. The models are based on data from Thomson Reuters Asset4 database. See appendix 8 for description of the variables.

CFP-E 1- CFP-E 3 are models based on the relationship between direct environmental exposure and corporate financial performance where Ksi1 is a latent construct for the direct environmental exposure. Eta1 is a latent construct for profitability and eta2 is a latent construct for growth. CFP-ESG 1-4 are models based on the relationship between environmental, social and governance factors and growth and profitability as dependent latent constructs. See appendix 8 for description and appendix 4 for the indicators. The relevance and reliability of the variables and models will be discussed in section 4.5.3. The description of the hypothesized latent constructs for the x-variables can be seen in table 4.15.

Model	Ksi	Description latent construct
CFP-E 1	Ksi1	Direct environmental exposure/impact
CFP-E 2	Ksi1	Direct environmental exposure/impact
CFP-E 3	Ksi1	Direct environmental exposure/impact
CFP-ESG 1	Ksi1	Direct environmental exposure/impact
CFP-ESG 2	Ksi1	Direct environmental exposure/impact
CFP-ESG 2	Ksi2	Corporate controversies
CFP-ESG 2	Ksi3	Board composition
CFP-ESG 3	Ksi1	Direct environmental exposure/impact
CFP-ESG 3	Ksi2	Corporate controversies
CFP-ESG 3	Ksi3	Human Rights compliance
CFP-ESG 3	Ksi4	Board composition
CFP-ESG 3	Ksi5	Audit independence
CFP-ESG 4	Ksi1	Direct environmental exposure/impact
CFP-ESG 4	Ksi2	Corporate controversies
CFP-ESG 4	Ksi3	Human Rights compliance

Table 4.15 Hypothesized latent constructs SEM-models

4.5.2 SEM-model equations

This section shows the equations used for the full SEM-models. $x_1 - x_{15}$ are the observed independent indicators for the latent variables $\xi_1 - \xi_5$. $y_1 - y_6$ are the dependent observed indicators for $\eta_1 - \eta_2$. $\delta_{i,j}$ and $\varepsilon_{i,j}$ represents the measurement errors for $x_{i,j}$ and $y_{i,j}$. The measurement error for $\eta_{i,j}$ is denoted $\zeta_{i,j}$. The relationship between the observed indicators and the latent factors are $\lambda^x_{i,j}$ and $\lambda^y_{i,j}$. $\gamma_{i,j}$ represents the relationship between $\xi_{i,j}$ and $\eta_{i,j}$ while $\beta_{i,j}$ represents the relationship between $\eta_{i,j}$. Full overview of the equations can be found in appendix 2.

4.5.3 Full SEM-models results and analysis

4.5.3.1 Estimated measurement models

Appendix 4 shows the indicators, standardized factor loading, covariance measurement error Θ_δ, standard error, t-values and R^2 for the observable proxies for the latent constructs. A high absolute value of the standardized factor loadings indicates that the proxies are a good indicator for the latent construct.

Appendix 4 shows that the standardized factor loadings are overall high for the environmental, social and governance proxies, but some are also very low or moderate. The table shows that most of the variables are significant on a 1% and 5% level, but some are also insignificant. Model CFP-ESG 3 could not estimate the t-values because the sample size was too small to estimate the asymptotic covariance matrix using RML. However, estimating the model with ML shows that most of the indicators for Ksi are significant on a 5% level, see appendix 5. We must interpret these t-values with caution due to non-normality. The table shows that the indicators of Ksi1 for all the models are good indicators for direct environmental exposure (Ksi1). However, the relevance of eco design products and environmental products can be questioned since they both have low standardized factor loadings. The indicators for Ksi2-Ksi5 have more moderate factor loadings compared to Ksi1.

Appendix 4 shows that all the observed indicators for profitability and growth are significant on a 1% level. All the standardized factor loadings are mostly above 0,5 except for 1-year net sales/revenue growth which ranges from 0,38-0,45. Overall, the selected indicators seem to be good indicators for the latent dependent constructs, eta1 and eta2, with significant and high standardized factor loadings.

4.5.3.2 Estimated structural model

Structual model						
Model	Indicator	Parameter	Standardized estimate	Standard error	t-value	R2
CFP-E 1	Ks1 -> eta1	γ1,1	-0,05	0,109	-0,483	
	eta1 -> eta2	β2,1	-0,12	0,015	0,408	
	eta1					0,003
	eta2					0,000
CFP-E 2	Ksi1 -> eta1	γ1,1	-0,05	0,108	-0,482	
	eta1					0,003
CFP-E 3	Ksi1 -> eta1	γ1,1	0,016	0,026	0,627	
	eta1					0,000
CFP-ESG 1	Ksi1-> eta1	γ1,1	-0,09	0,127	-0,713	
	Ksi2 -> eta1	γ1,2	0,09	0,098	-0,831	
	Ksi3 -> eta1	γ1,3	0,11	0,101	-1,38	
	Ksi1 -> eta2	γ2,1	-0,08	0,107	0,979	
	ksi2 -> eta2	γ2,2	0,11	0,099	1,254	
	Ksi3 -> eta2	γ2,3	-0,2	0,092	-1,915	
	eta1					0,034
	eta2					0,058
CFP-ESG 2	Ksi1-> eta1	γ1,1	-0,09	0,127	-0,684	
	Ksi2 -> eta1	γ1,2	-0,08	0,098	-0,816	
	Ksi3 -> eta1	γ1,3	0,12	0,201	-1,479	
	Ksi1 -> eta2	γ2,1	-0,08	0,104	0,809	
	ksi2 -> eta2	γ2,2	0,12	0,102	1,137	
	Ksi3 -> eta2	γ2,3	0,5	0,378	1,316	
	eta1					0,103
	eta2					0,268
CFP-ESG 3	Ksi1 -> eta1	γ1,1	-0,05	-	-	-
	Ksi2 -> eta1	γ1,2	0,03	-	-	-
	Ksi3 -> eta1	γ1,3	-0,47	-	-	-
	ksi4 -> eta1	γ1,4	-0,46	-	-	-
	ksi5 -> eta1	γ1,5	0,02	-	-	-
	Ksi1 -> eta2	γ2,1	0,11	-	-	-
	Ksi2 -> eta2	γ2,2	0,03	-	-	-
	Ksi3 -> eta2	γ2,3	-0,16	-	-	-
	ksi4 -> eta2	γ2,4	0,51	-	-	-
	ksi5 -> eta2	γ2,5	-0,02	-	-	-
	eta1					0,433
	eta2					0,306
CFP-ESG 4	Ksi1 -> eta1	γ1,1	-0,05	0,109	-0,466	
	Ksi2 -> eta1	γ1,2	-0,04	0,093	-0,463	
	Ksi3 -> eta1	γ1,3	-0,46	0,415	-1,102	
	Ksi1 -> eta2	γ2,1	0,1	0,108	0,954	
	Ksi2 -> eta2	γ2,2	0,1	0,093	1,101	
	Ksi3 -> eta2	γ2,3	-0,293	0,267	-1,099	
	eta1-> eta2	β2,1	-0,289	0,139	-2,085	
	eta1					0,213
	eta2					0,119

Table 4.16 Estimated structural model for full ESG-CFP models

Table 4.16 shows the estimated structural models. The table shows the parameters, standardized estimate, standard error, t-value and R^2. The structural parameters indicate one-unit change in the explanatory variable will lead to a change in the dependent variable with the value of the estimated structural parameter. The table show that almost none of the estimated structural parameters are significant on a 10% level, but some are. The standardized estimates have both negative and positive effect on profitability and growth, but are not significant. The non-significance could be a result based on a poor measurement instrument. However, Ksi1 has a weak negative effect on profitability for all the models. Ksi1 have both a non-significant positive and weak effect on growth. Ksi2-Ksi5 have different signs for the different models.

The R^2 for all the CFP-E models are close to 0 and indicates an unreliable model for the structural model, but the indicators for Ksi seem reliable. The CFP-ESG models have higher R^2 for the equations, but almost none of parameters are significant on a 10% level, and this makes them unreliable. Table 4.16 shows that 21,3% of the variance in profitability and 11,9% of the variance in growth are explained by the hypothesized latent constructs direct environmental exposure/impact, corporate controversies and human rights compliance.

4.5.3.3 Goodness of fit and reliability

Goodness of fit indices	CFP-E 1	P-value	CFP-E 2	P-value	CFP-E 3	P-value	CFP-ESG 1	P-value	CFP-ESG 2	P-value	CFP-ESG 3	P-value	CFP-ESG 4	P-value
C1	49,491	0,1705 **	17,797	0,4961 **	17,648	0,0395 *	71,911	0,4144 **	83,689	0,4582 **	198,424	0,021 *	97,236	0,4456 **
C3	61,081	0,0225 *	23,145	0,0822 *	26,865	0,0015	83,349	0,1316 **	94,893	0,1753 **	204,399	0,0102 *	107,949	0,1903 **
RMSEA	0,037	0,705 **	0	0,807 **	0,0478	0,496 **	0,0158	0,886 **	0,00857	0,93 **	0,0461	0,607 **	0,0107	0,941 **
NFI	0,939 *		0,974 **		0,969 **		0,912 *		0,903 *		0,834		0,896	
NNFI	0,971 **		0,991 **		0,965 *		0,98 **		0,983 **		0,949		0,984 **	
PNFI	0,7		0,626		0,582		0,701		0,713		0,703		0,717	
CFI	0,979 **		0,994 **		0,979 **		0,984 **		0,986 **		0,957 *		0,987 **	
IFI	0,979		0,994		0,979		0,985		0,987		0,959		0,987	
RFI	0,918		0,959		0,949		0,885		0,877		0,803		0,87	
Critical N	160,502		226,566		339,734		132,333		137,771		113,073		137,066	
RMR	6,954		0,829		0,892		5,194		7,488		22,903		4,677	
SRMR	0,064 *		0,0275 **		0,0331 **		0,0591 *		0,0677 *		0,0859 *		0,0722 *	
GFI	0,948 *		0,973 **		0,986 **		0,919 *		0,911 *		0,861		0,909 *	
AGFI	0,916 **		0,947 **		0,968 **		0,879 *		0,871 *		0,817		0,871 *	
PGFI	0,589		0,487		0,423		0,879		0,63		0,656		0,642	
AIC	6177,068		3218,519		8981,365		4253,116		5865,388		6934,164		3863,836	
BIC	6252,5		3272,83		9029,876		4347,632		5966,301		7070,534		3972,931	
Test for multivariate normality	Test statistic	P-value	Test statistic	P-value	Test statistic	P-value	Test statistic	P-value	Test statistic	P-value	Test statistic	P-value	Test statistic	P-value
Skewness	Z=43,730	0	Z=38,852	0	Z=52,459	0	Z=38,797	0	-	-	-	-	-	-
Kurtosis	Z=15,582	0	Z=14,672	0	Z=19,658	0	Z=12,019	0	-	-	-	-	-	-
Skewness & kurtosis	Chi2=2068,72	0	Chi2=1655,60	0	Chi2=3142,57	0	Chi2=1649,64	0	-	-	-	-	-	-

**) Good fit
*) Acceptable fit

Table 4.17 Goodness of fit indices for the full SEM-models

Table 4.17 shows the goodness of fit indices based on Schermelleh-Engel et al. (2003). The table shows that all the models reject the null hypothesis of multivariate normal distribution, but model CFP-E 3 and CFP-E 4 do not have a large enough sample size to estimate the asymptotic covariance matrix that contains information about skewness and kurtosis. However, the models are estimated with RML based on the assumption that they are not multivariate normal distributed.

Table 4.17 also shows that the difference between C1 and C3 is small, which indicates that the non-normality is not a problem, but C3 is the more reliable Chi-square. However, we have a small sample and the test is more often rejected in a large sample than a small sample (Sharma et al. 2005; Schermelleh-Engel et al., 2003). RMSEA also indicates good or acceptable fit for all the models. The incremental indices, NFI, NNFI and CFI, indicate good and acceptable fit for the models, except for CFP-ESG 3. If the incremental indices deviate a lot from GFI and RFI it is an indication of disturbance in the data, but it does not appear to be a problem. However, model CFP-ESG 3 does not appear to have a good fit based on the incremental indices. The models with Critical N < 200 imply poor fit. Appendix 4 shows that the R^2 for the observed x- and y-variables have mostly high values and above 0,5, except for some x-variables close to 0.

Measures of reliability				
Model	Latent factor	Composite reliability	Average Variance extracted	Cronbach's Alpha
CFP-E 1	Ksi1	0,74	0,50	0,63
	eta1	0,93	0,82	0,80
	eta2	0,68	0,44	0,82
CFP-E 2	Ksi1	0,74	0,50	0,63
	eta1	0,93	0,82	0,80
CFP-E 3	Ksi1	0,80	0,57	0,74
	eta1	0,60	0,52	0,54
CFP-ESG 1	Ksi1	0,78	0,53	0,66
	Ksi2	0,89	0,80	0,88
	Ksi3	-	-	-
	eta1	0,94	0,83	0,92
	eta2	0,74	0,50	0,66
CFP-ESG 2	Ksi1	0,78	0,52	0,66
	Ksi2	0,88	0,79	0,88
	Ksi3	0,00	0,02	0,46
	eta1	0,94	0,84	0,92
	eta2	0,73	0,49	0,66
CFP-ESG 3	Ksi1	0,78	0,54	0,66
	Ksi2	0,90	0,81	0,88
	Ksi3	0,53	0,32	0,44
	Ksi4	0,01	0,02	0,46
	Ksi5	0,88	0,78	0,82
	eta1	0,94	0,84	0,92
	eta2	0,73	0,50	0,66
CFP-ESG 4	Ksi1	0,78	0,53	0,66
	Ksi2	0,89	0,79	0,88
	Ksi3	0,53	0,32	0,44
	eta1	0,94	0,63	0,92
	eta2	0,73	0,50	0,66

Table 4.18 Reliability measures full SEM-models

Table 4.18 shows that not all the latent factors have an acceptable value. Ksi3 model CFP-ESG 2, Ksi3 and Ksi4 model CFP-ESG 3 and Ksi3 model CFP-ESG 4 have CR values below 0,6. Most of the latent factors have an AVE-value above 0,50 and some have more moderate values ranging from 0,3-0,4. Ksi4 model CFP-ESG 3 have a value of 0,02 and indicates a non-acceptable fit and poor reliability. The lower boundary of an acceptable Cronbach's alpha is 0,7 and the table shows that some latent factors have a value above 0,7. The low CA values indicate low internal reliability and is caused by low correlation between the indicators.

4.5.3.4 Discussion of results

The simplest models, CFP-E 1 and CFP-E 3, analyzing the direct relationship between direct environmental exposure, profitability and growth, indicate that the selected indicators for Ksi have a good fit, based on the discussion in 4.5.3.1- 4.5.3.3. This holds, even though environmental products and eco design products have low standardized factor loadings. Both models imply that direct environmental exposure has a very low negative effect on profitability, but both estimates are not significant on a 5% level. Model CFP-E 1 implies that profitability has a low negative effect on growth, but only significant on a 10% level. Both models have estimated the covariance between the measurement errors for eco design and environmental products due to both indicators being closely related. Even though the models show god fit, the CPF-E are models unreliable due to low R^2 and the insignificant structural model.

When the sample size increases the selected indicators for growth and profitability seem to be less reliable. Model CFP-E 3 had to drop to some indicators due to a negative measurement error close to zero, and one indicator is fixed to 1 because of this. The model adapts to this fixation. Model CFP-E 3 may indicate that the results in model CFP-E 1 and CFP-E 2 is caused by a small sample and industry bias or a poorly specified model.

Model CFP-ESG 1 to CFP-ESG 4 are very similar models, but with some different indicators for the independent observed x-variables. Many dichotomous variables have been omitted for the effective sample size due to having only one value, and limits what is possible to test. The ESG data quality makes it hard to model the relationship. Based on the discussion in 4.5.3.1- 4.5.3.3 the CFP-ESG models have an acceptable fit, except for CFP-ESG 3, but almost no significant factor loadings for the structural model. This may indicate a poorly specified model and/or omitted relevant indicators. Overall, the structural model is unreliable.

Based on our results we can conclude our SEM-model hypotheses from 3.2.2. H1S is rejected for all the models. Even though we have mostly reliable indicators for the environmental factor, no model has a significant effect on profitability and growth. H2S is also rejected for all the models. The SEM-models do not show any significant positive or negative effect on growth and profitability for the social factors. H3S has not been tested properly because including governance data in the full SEM-models cause disturbance in the model due to a low sample size and many omitted variables. Model CFP-ESG3 is an example of this. Model CFP-ESG2 has tested some governance variables, but no significant effect on profitability and growth. H4S is confirmed by model CFP-ESG4 and shows a negative significant effect, but due to data we must question the reliability of the model.

However, the relationship between ESG and CFP is complex. Fatemi, Glaum and Kaiser (2018) created a SEM-model using ESG data on an overall level and some (few) disaggregated variables for the relationship between ESG and CFP using only market value as a proxy for CFP. They found that ESG strengths increase firm value, but when ESG disclosure interacted with ESG strengths or weakness, it weakens the positive valuation effect of ESG strengths. They argue that stock markets may interpret stepped-up disclosure as an attempt to justify overinvestment in ESG. They also argue that disclosure may help firms legitimate their behavior by explaining to investors the appropriateness of their operations and ESG policies or that they have made commitments to change their operations and overcome weakness. Using the same model for the individual E, S and G score they find that environmental strengths increases firm value and environmental weakness decrease it. Social and governance weakness decreased the firm value, but social and governance strengths had no effect. Based on this they argue that investors may discriminate strongly among the different dimensions of ESG.

One could argue that market value is influenced by profitability over time when operationalized as a quality factor like NBIM (2015) does. This raises the question how environmental, social and governance data on a disaggregated level influence profitability. The environmental data, for the CFP-E and CFP-ESG, are more related to core operation than social and governance, hence using data measuring pollution and resource use. Our experience with the governance and social data is that the companies are “too homogenous”. Both the European and U.S companies have a very small variation when it comes to reporting governance and social policies. As an example, companies report on policies they have already implemented and do not report policies they have not implemented. For many variables this results in a situation with dichotomous variables having only one value and is omitted from the model. Fatemi, Glaum and Kaiser (2018) argue that this is caused by the Security and Exchange Commission (SEC) and other financial regulatory authorities, and that the social and environmental data are mostly voluntary and might be difficult for investors to verify.

4.5.3.5 Validity

Construct validity means that we measure the theoretical concept we are trying to measure (Ringdal, 2018). It is not easy to measure validity compared to reliability, and there is no method to measure validity. The relationship between CFP and ESG is very complex and in practice would consist of more latent factors that are included in the SEM-models presented here. A lot of ESG data are available in the Thomson Reuters Asset4 database, but most of them are dichotomous variables and are not that informative about direct corporate operations. However, the data must be standardized to be used as indicators. The selected indicators are chosen on an explorative basis, based on the EFA and CFA, because we found very little empirical research on ESG-CFP SEM-models. The models presented are trying to analyze the relationship between CFP and ESG using data that are somewhat directly linked to core operations. Reliability is a prerequisite for validity and the mixed reliability, low t-values and R^2 for the structural model, do not imply a valid model.

4.6 Panel data regression

To compensate for the SEM-models' weakness in relationship to stock market corporate financial performance, a panel data regression model has been constructed to analyze the relationship between CFP and ESG. The purpose of the regression models is to analyze whether ESG variables, among other variables, can explain yearly stock returns. The models are tested for random or fixed effects, heteroskedasticity and autocorrelation. See appendix 3 for notation for the heteroskedasticity and autocorrelation tests. All the regression models have yearly stock return as a dependent variable. Following the analysis of the panel data regression results, the models will be compared, and reliability will be discussed.

4.6.1 Analysis panel data regression results

Hausman test	Test statistic	Df.	P-value
Model 1	Chi2= 52,02	14	0,000
Model 2	Chi2=125,41	13	0,000
Model 3	Chi2= 206,42	13	0,000
Model 4	Chi2= 138,11	14	0,000
Model 5	Chi2=141,2	16	0,000

Table 4.19 Results Hausman test for fixed or random effects

Table 4.19 shows that the fixed effects regression model is the correct regression model compared to a random effects model. Test specification for the Hausman test is shown in chapter 3.2.4.2**.** The results are as expected because one would assume that there are individual differences between sectors, within sectors and across time. All the models are significant at a 5% level.

The correlation matrix shown in appendix 9 shows the pairwise correlation coefficient, number of observations and the p-value. The correlation matrix shows that there is a relative strong positive significant linear covariation for ROA, ROIC and ROE, something which is expected. The correlation between the scores for environment, social and governance ranges from 0,3 to 0,7 and are significant at a 5% level. The controversies score is negatively correlated, as expected, with the environmental, social and governance score. For the rest of the variables there is no strong positive or negative correlation between the variables and many pairwise correlations are significant and very close to zero. Using too many indicators of CFP has been avoided because accounting return measures are often strongly correlated. This implies that multicollinearity is most likely not a problem for the regression models.

Wooldridge's test for autocorrelation in panel data	Test statistic	Degree of freedom	P-value
Model 1 and 2	F=2,851	F(1,914)	0,0916
Model 3 and 4	F=3,112	F(1,913)	0,0781
Model 5	F=3,017	F(1,942)	0,0827

Table 4.20 Results Wooldridge's test for autocorrelation in panel data

Table 4.20 shows the test for autocorrelation in panel data models. The tests are done by using Wooldridge's test for autocorrelation for each model in STATA 16. The test results indicate that autocorrelation is not necessarily a problem for the panel data regression models and the null hypothesis of no first-order autocorrelation is not rejected using a 5% significance level.

Breusch-Pagan Lagrange Multiplier panel heteroskedasticity test	**Test statistic**	**Degree of freedom**	**P-value**
Model 1 and 2	Chi2= 916, 807	Df= 8	0,000
Model 3 and 4	Chi2= 927,12	Df. =8	0,000
Model 5	Chi2= 919,09	Df. =8	0,000

Table 4.21 Results Breusch-Pagan Lagrange Multiplier panel heteroskedasticity test

Table 4.21 shows that heteroskedasticity is a problem in all the panel data regression models, based on the test developed by Shehata (2012), and the null hypothesis of panel homoscedasticity is rejected for all the models. The variables in each model are transformed to reduce the effect of outliers, but this is not enough for the tests to not reject the null hypothesis. To reduce the problem of heteroskedasticity we have estimated the standard heteroskedasticity corrected (robust) standard errors. The robust standard errors do not change the regression coefficients, only the standard errors.

Modified Wald test for groupwise heteroskedasticity in fixed effect regression	Test statistic	P-value
Model 1	Chi2= 5,3*10^34	0,000
Model 2	Chi2=8,4*10^9	0,000
Model 3	Chi2= 1,3*10^16	0,000
Model 4	Chi2= 7,0*10^9	0,000
Model 5	Chi2= 2,6*10^10	0,000

Table 4.22 Results Modified Wald test for groupwise heteroskedasticity in fixed effect regression

Table 4.22 shows the modified Wald test for heteroskedasticity and shows that groupwise heteroskedasticity is a problem for all the models. However, the test does not work well in panel data where N is large and t is small, so the test must be interpreted with caution (Greene, 2000). For all our panel data regression models N=993 and t=9 and therefore the validity of the test must be questioned. Both tests for heteroskedasticity justify using robust standard errors, and for some variables the t-values drop significantly and indicates that heteroskedasticity is a problem before robust estimation.

4.6.2 Fixed effects regression model 1

	Model 1				
Variable	Parameter estimate $\hat{\beta}$	t-value	Robust standard error	P-value	
ESG Score	-0,1969	-3,41	0,0576	0,001	**
ROE	0,004	0,71	0,0056	0,479	
ROA	0,1711	1,44	0,1191	0,151	
Natural log capital expenditure/assets	-2,0949	-2,29	0,9145	0,022	**
Natural log assets	-4,6733	-3,37	1,3885	0,001	**
Natural log Debt/assets	0,6934	0,86	0,8041	0,389	
2011	-26,3894	-17,96	1,4691	0,000	**
2012	-3,2054	-2,07	1,5477	0,039	**
2013	10,2034	5,49	1,8596	0,000	**
2014	-9,2455	-5,94	1,5556	0,000	**
2015	13,7309	-8,8	1,5607	0,000	**
2016	-11,3618	-6,97	1,6299	0,000	**
2017	-0,453	-0,28	1,6023	0,777	
2018	-28,498	-17,14	1,6623	0,000	**
Constant	109,7123	4,9	22,4035	0,000	**
Rho	0,2009				
F-test for rho				0,201	
F-test for overall significance		96,02		0,000	**
R-square:					
within	0,1862				
between	0,0613				
overall	0,1493				

**) Significant on a 5% level

Table 4.23 Results estimated fixed effects regression model 1

Table 4.23 shows that the ESG score is a significant explanatory variable while controlling for ROA, ROE, capital expenditure/assets, size, and debt/assets. The model has a time variable, year, where the reference variable is year 2010 and the model must be interpreted in relationship to 2010. However, we see that ROA and ROE are not significant explanatory variables. Log capital expenditure in percent of assets and log assets are both control variables and control for capital intensive companies and size. Log debt/assets is also a control variable and controls for companies in financial distress but are not significant. The F-test for overall significance rejects the null hypothesis of all regression coefficients equal to zero. ρ (rho) shows that 20% of the variance is explained by entity (company) specific differences but is not significant.

For model 1 it is expected that when the ESG score increases by one, the stock returns decrease 0,17% given that all other variables are hold constant. When the capital expenditure/asset ratio increases by 1% the stock return is expected to decrease roughly by –0,0209% given that all other variables are hold constant. When assets increase by 1% the stock return is expected to decrease by –0,0467% given that all other variables are hold constant. Most of the time variables are significant except 2017.

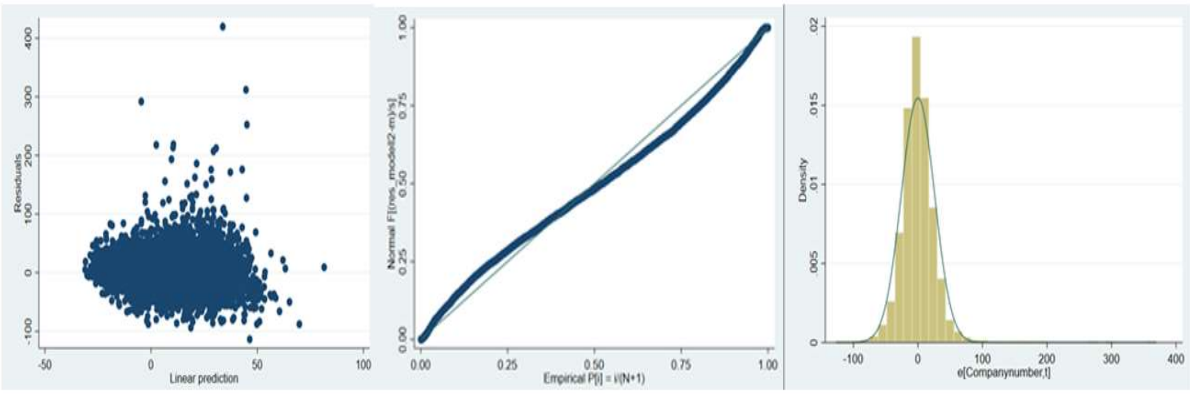


Figure 4.2 Residual diagnostics charts model 1: Residual-plot, probability-plot and histogram of residual distribution compared to the normal distribution.

Figure 4.2 shows the residual plot, probability plot and the distribution of the residuals compared to the normal distribution. The residual-plot shows that the residuals are not randomly spread, but they show a trend, and this indicates a problem with heteroskedasticity. The probability-plot compares the residuals to the normal distribution and shows that the residuals have a greater variance that the normal distribution (Hammervold, 2020). The line is S-shaped and indicates that the residuals are non-normal distributed. The histogram shows that the residuals have a kurtosis greater than the normal distribution, but the distribution is not left or right skewed.

4.6.3 Fixed effects regression model 2

Variable	Model 2 Parameter estimate $\hat{\beta}$	t-value	Robust standard error	P-value	
ESG Score	-0,097	-1,570	0,062	0,117	
Lag(1) ROE	0,023	2,630	0,009	0,009	**
Lag(1) ROA	-0,696	-5,380	0,129	0,000	**
Natural log capital expenditure/assets	-1,897	-2,030	0,933	0,000	**
Natural log assets	-4,307	-2,950	1,459	0,042	**
Natural log Debt/assets	-0,733	-1,010	0,727	0,003	**
2012	22,768	15,920	1,431	0,314	
2013	35,716	21,200	1,685	0,000	**
2014	16,647	14,090	1,181	0,000	**
2015	11,575	8,480	1,366	0,000	**
2016	13,220	9,350	1,414	0,000	**
2017	24,389	16,510	1,477	0,000	**
2018	-3,363	-2,420	1,392	0,000	**
Constant	81,922	3,380	24,202	0,001	**
Rho	0,252				
F-test rho		1,170		0,0004	**
F-test overall significance		90,130		0,000	**
R-square:					
within	0,2004				
between	0,0004				
overall	0,1216				
*) Significant on a 5% level					

Table 4.24 Results estimated fixed effects regression model 2

Table 4.24 shows that model 2 has the same variables as model 1, but a time (year) lag for ROE and ROA has been used. The model is interpreted in the same way as model 1 except for ROE and ROA. The reasoning for this model is stock markets' reaction to annual reports/income statements. The question is wheter the previous yearly, t-1, ROA and ROE influence stock return given the same control variables. Year 2011 is the reference variable for the time variable.

The ESG score is no longer a significant explanatory variable in model 2, and the lagged variables and the control variables are now significant on a 5% level. The variance explained by the entity specific differences is 25% and the overall F-test is significant on a 5% level. The signs for the time variables are all positive, except for 2018, and significant on a 5% level. When ROE increases by one percent the stock return is expected to increase 0,023%. When ROA increases by one percent the stock return is expected to decline 0,7% given all other variables are hold constant. Figure 4.3 shows that the residual diagnostics are similar to model 1.

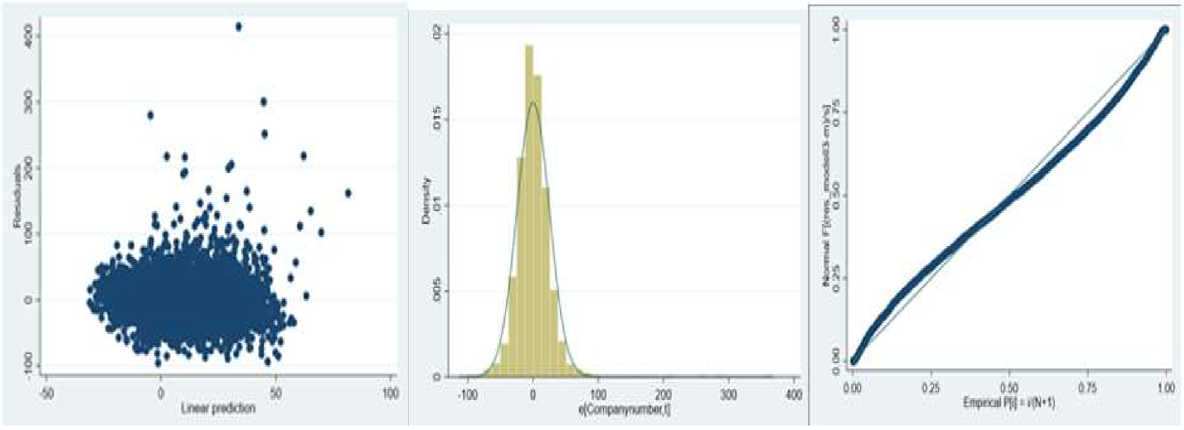


Figure 4.3 Residual diagnostics charts model 2: Residual-plot, probability-plot and histogram of residual distribution compared to the normal distribution.

4.6.4 Fixed effects regression model 3

	Model 3				
Variable	Parameter estimate $\hat{\beta}$	t-value	standard error	P-value	
Natural log ESG score	-6,456	-2,000	3,229	0,046	**
Lag(1) ROE	0,027	2,670	0,010	0,008	**
Lag(1) ROA	-0,693	-5,380	0,129	0,000	**
Natural log Capital expenditure/assets	-2,085	-2,270	0,919	0,024	**
Natural log assets	-4,109	-2,850	1,443	0,004	**
Natural log Debt/assets	-0,835	-1,130	0,738	0,258	
Natural log Lag(1) Cash flow operation/sales	-1,972	-1,830	1,076	0,067	*
2012	22,340	15,770	1,417	0,000	**
2013	35,039	21,030	1,666	0,000	**
2014	16,586	13,800	1,202	0,000	**
2015	11,313	8,320	1,360	0,000	**
2016	13,615	9,540	1,428	0,000	**
2017	24,417	16,570	1,474	0,000	**
2018	-3,328	-2,400	1,387	0,067	*
Constant	105,0765	3,86	27,2365	0,000	**
Rho	0,257				
F-test rho		1,190		0,001	**
F-test		81,180		0,000	**
R-square:					
within	0,2063				
between	0,002				
overall	0,1237				
*) Significant 5% level					
**) Significant 10% level					

Table 4.25 Results estimated fixed effects regression model 3

Model 3 uses the same variables as model 2 except for log of ESG score. Model 3 shows that all the variables are significant on a 5% level except for debt/assets and natural log lag (1) cash flow operations/sales, using lag of the CFP indicators and the log of the ESG score.

In this model an increase by 1% in the ESG score would give an expected decline in the stock return of 0,006% controlled for the other variables. An increase in ROA the previous year by 1% would give an expected decrease in stock return by 0,693% controlled for the other variables. Further there is a significant positive effect for the specific years compared to 2011, except for the effect of 2018, which is negative and only significant on the 10% level.

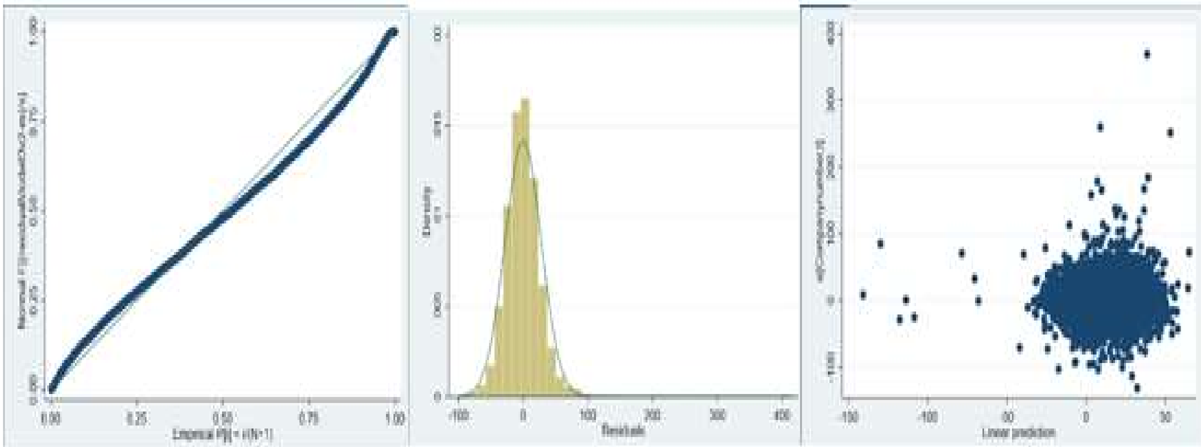


Figure 4.4 Residual diagnostics charts model 3: Residual-plot, probability-plot and histogram of residual distribution compared to the normal distribution.

However, figure 4.4 shows that the residual plot contains a trend and the residuals are not randomly spread. The probability plot shows that the residuals have a greater variance than the normal distribution. The histogram shows that the residuals have a higher kurtosis than the normal distribution.

4.6.5 Fixed effects regression model 4

	Model 4				
Variable	Parameter estimate $\hat{\beta}$	t-value	Robust standard error	P-value	
Natural log Lag(1) ESG score	-5,379	-1,750	3,076	0,081	*
Lag(1) ROE	0,022	2,510	0,009	0,012	**
Lag (1) ROA	-0,679	-5,320	0,128	0,000	**
Natural log capital expenditure/assets	-1,937	-2,070	0,937	0,039	**
Natural log debt/assets	-0,709	-0,970	0,731	0,332	
Natural log assets	-4,235	-2,880	1,471	0,004	**
Lag (1) Cash flow operations/sales	-0,004	-5,580	0,001	0,000	**
2012	22,783	14,780	1,444	0,000	**
2013	35,599	21,560	1,651	0,000	**
2014	16,537	13,890	1,191	0,000	**
2015	11,028	8,080	1,365	0,000	**
2016	13,028	9,710	1,391	0,000	**
2017	24,542	16,550	1,483	0,000	**
2018	-2,874	-2,150	1,336	0,032	**
Constant	96,210	3,910	27,390	0,000	**
F-test overall significance		99,430		0,000	**
Rho	0,255				
F-test rho		1,14		0,003	**
F-test for overall significance		99,43		0,000	**
R-square:					
within	0,2073				
between	0,0008				
overall	0,127				
**) Significant 5% level					
*) Significant 10% level					

Table 4.26 Results estimated fixed effects regression model 4

Table 4.26 shows that model 4 has the same variables as model 3 except cash flow operation/net sales and lag (1) of log ESG score. However, we see that using a lagged ESG score the variable is no longer significant on a 5% level and ρ is lower compared to model 3. Figure 4.5 shows that model 4 has very similar residual properties as model 1,2 and 3, indicating that heteroskedasticity is a problem, but the standard errors are heteroskedasticity corrected for all the models. The signs for the time variables are unchanged and significant. All the control variables are also significant except debt/assets. Model 4 indicates that the ESG score for the previous year might not be as good as an indicator compared to the ESG score for the current year.

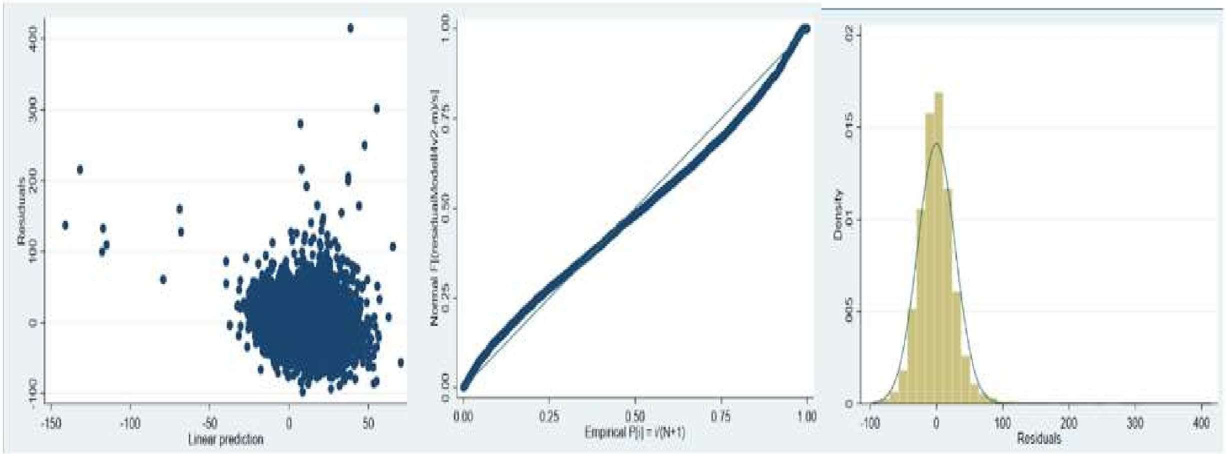


Figure 4.5 Residual diagnostics charts model 4: Residual-plot, probability-plot and histogram of residual distribution compared to the normal distribution.

4.6.6 Fixed effects regression model 5

	Model 5				
Variable	Parameter estimate $\hat{\beta}$	t-value	Robust standard error	P-value	
Natural log controversies score	0,196	0,520	0,380	0,606	
Natural log environmental score	-5,132	-2,260	2,267	0,024	**
Natural log social score	4,822	1,850	2,608	0,065	*
Natural log governance score	-5,426	-2,970	1,826	0,003	**
Lag(1) ROE	0,022	2,570	0,009	0,010	**
Lag(1) ROA	-0,699	-5,380	0,130	0,000	**
Natural log assets	-4,258	-2,950	1,442	0,003	**
Natural log Debt/assets	-0,729	-1,020	0,717	0,310	
Natural log capital expenditure/assets	-1,992	-2,160	0,923	0,031	**
2012	22,849	15,950	1,433	0,000	**
2013	35,827	21,220	1,688	0,000	**
2014	16,841	14,180	1,188	0,000	**
2015	11,913	8,610	1,383	0,000	**
2016	13,774	9,780	1,408	0,000	**
2017	24,891	16,800	1,481	0,000	**
2018	-2,568	-1,840	1,394	0,066	*
Constant	97,033	3,660	26,487	0,000	**
Rho	0,253				
F-test rho		1,18		0,0003	**
F-test overall significance		73,61		0,0000	**
R-square:					
within	0,2067				
between	0,0006				
overall	0,1229				
**) Significant 5% level					
*) Significant 10% level					

Table 4.27 Results estimated fixed effects regression model 5

Table 4.27 shows that model 5 uses the scores that the combined ESG score consists of, namely the environmental, social, governance and controversies score while controlling for size, leverage, and capital intensity. The model shows that the environmental and governance scores are significant on a 5% level and the social score on a 10% level. Neither leverage nor controversies scores are significant on a 10% level. Figure 4.6 indicates that heteroscedasticity is also a problem for model 5 and the residuals are not randomly spread and have very similar residual-plots compared to model 1,2,3 and 4.

The regression coefficients indicate that the different scores may have a weak relationship to yearly stock returns. However, the environmental and the governance scores imply that by a one percent change in the scores, the expected stock return will decline respectively by 0,05132% and 0,05426% given that all other variables are constant. The controversies sign is positive implying that companies involved in different negative media stories are perceived as positive, and have a weak positive effect on stock returns, but are not significant on a 10% level.

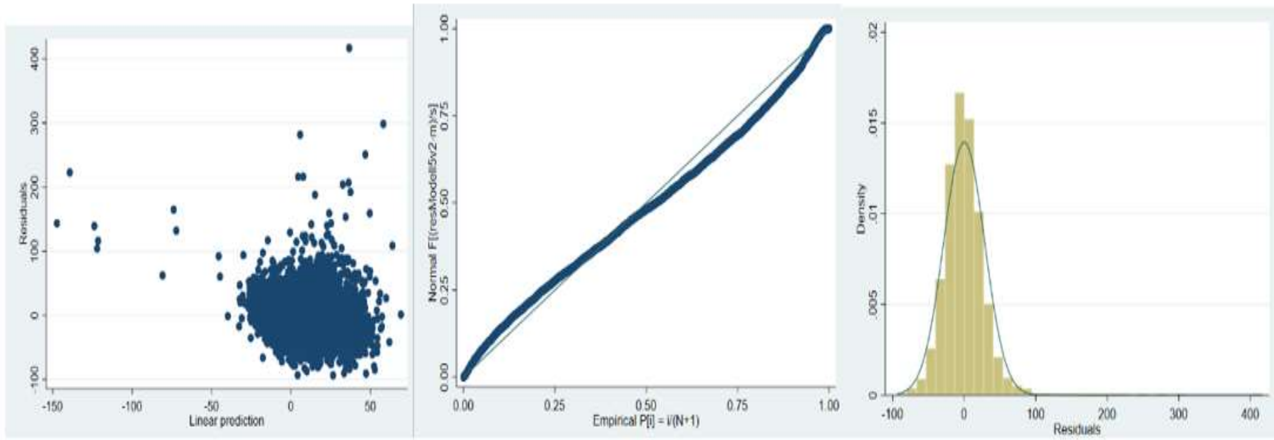


Figure 4.6 Residual diagnostics charts model 5: Residual-plot, probability-plot and histogram of residual distribution compared to the normal distribution.

4.6.7 Model comparison and discussion of results

Model 3 implies that the ESG score is a significant explanatory variable if there is a one-year lag on the indicators of CFP. Model 1 implies that the ESG score is a significant explanatory variable, but the indicators of CFP are not. The different models show that the level of significance might be dependent on model specification and transformations of variables. Model 4 might imply that ROA and ROE are more relevant for explaining the stock returns given the assumption that the annual income statement might have some effect. The question is if this information is already taken into consideration and in addition the magnitude of earnings surprises.

Model 5 implies that the environmental and governance scores have a negative effect on the stock return. Of course, it is not given that the more environmentally friendly company is the more profitable one, it might be the opposite. Governance might have a more indirect effect on stock return and is not that strongly related to the core operation as e.g. the environmental score. The social score has a positive sign but might also have a more indirect effect. Another question is the relevance of the control variables for all the models. Log assets is a proxy for size based on the assumption that small companies may have more growth potential and/or large companies might be less volatile. Log debt/assets is a proxy for companies in financial distress based on the assumption that those companies might be more volatile. However, some sectors like industry and energy are capital intensive and may require higher leverage compared to other sectors. Therefore, capital intensity is also controlled for. There might also be differences within each sector that should be controlled.

Variable	Sample size (#panels)	Model	Restriction	P-value
Natural log ESG score	N=753	3	ESG score >= 62 & <= 100	0,673
Lag (1) ROE	N=753	3	ESG score >= 62 & <= 100	0,155
Lag (1) ROA	N=753	3	ESG score >= 62 & <= 100	0,000
Rho=0,4694	N=753	3	ESG score >= 62 & <= 100	0,000
Natural log ESG score	N=603	3	ESG score <= 62	0,179
Lag (1) ROE	N=603	3	ESG score <= 62	0,199
Lag (1) ROA	N=603	3	ESG score <= 62	0,001
Rho=0,3634	N=603	3	ESG score <= 62	0.0043

Table 4.28 Restriction ESG fixed effects regression models.

The reliability of the models and their robustness must also be questioned. The amount of variance explained by the entity specific differences is approximately 20% and may indicate omitted variables or a wrongly specified model. Wooldridge's test for autocorrelation cannot be used on the models with the lagged variables, but one can assume that autocorrelation is most likely not a problem because the models use yearly data from 2010 to 2018. The tests for autocorrelation are performed without lags. However, table 4.28 shows that when the values of the ESG score are restricted, the variable is no longer significant even with a large sample size, but ROA is. The average ESG score is 62 and figure 4.1 shows that the distribution is left skewed. However, ρ increased to approximately 35-46% and is significant. The F-test for overall significance is also significant on a 5% level. This may indicate that the ESG score is not a relevant explanatory variable for companies above average ESG score, and the results may be dependent on a large sample size. The same goes for the ESG score below average. However, the reliability of this approach to check robustness must be questioned because parts of the dataset are ignored.

Based on the regression models that have been tested the hypotheses can be concluded. H1P is partially confirmed and partially rejected. The results show that the ESG score is a significant explanatory variable, but not for all the models. It is significant when ROA and ROE are not lagged and become significant again when the natural logarithm has been applied to it. The hypothesis of the ESG score having a positive sign is rejected since all the models show negative values. H2P is rejected because there are no statistically significant results for ROA and ROE unless they are lagged. H3P is rejected for lagged ROA, since there are only significant negative effects shown to emerge. It is however a strong indication for lagged ROE having a positive relationship since it is positive and significant in three of the four models where it is included. H4P is rejected for controversies since it is not significant. It is confirmed for the environmental and governance pillar scores since they both exhibit a negative effect. There is some support for the social dimension, but only significant on the 10% level.

There are some studies that share similar results as regression model 1-5. Ahlklo and Lind (2019) also use a regression model trying to explain stock return for Nordic markets and find that the ESG, environmental and social scores are negative and insignificant. The governance score is slightly positive, but insignificant. Drange & Nath (2019) created a panel fixed effect regression model of 3450 companies to analyze the effect of indicators of CFP on the ESG, E, S and G scores. They find that ROA has a negative effect on all the different scores using different weights. The results in model 1-5 are consistent compared to Nollet, Filis & Mitrokostas (2016) who find that the ESG disclosure score has a negative impact on stock return, however they do not find any statistical significance for this.

4.7 Time series approach using portfolios

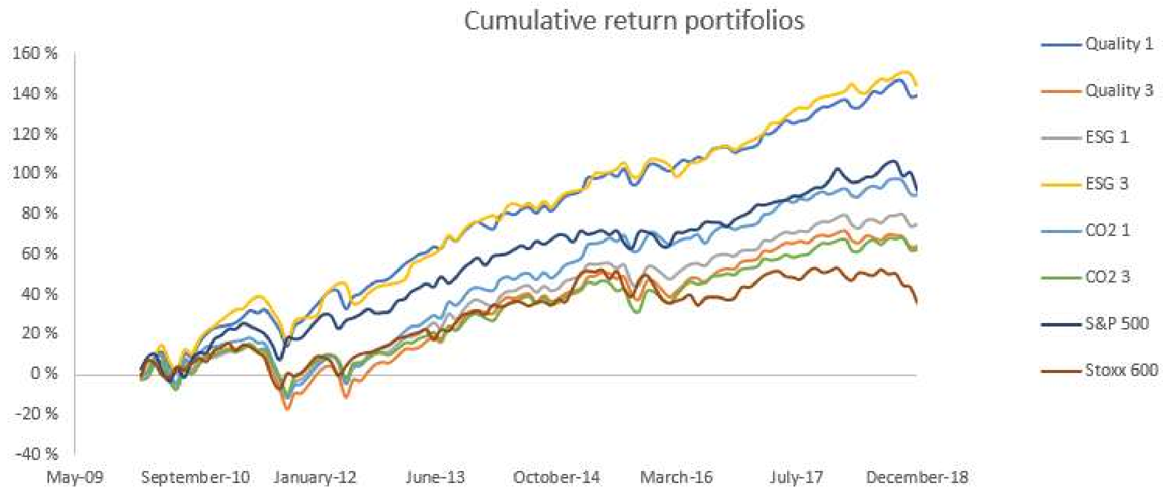


Figure 4.7 Cumulative return portfolios described in chapter 3.

Figure 4.7 shows the cumulative return of the stock portfolios described in chapter 3.3.1. The figure shows that the portfolio of bottom 33% (ESG 3) of ESG score has the highest cumulative return of 143% at the end of 2018. ESG 3 is closely followed by the Quality 1 portfolio. The figure also shows that the top 33% of ESG scores have a lower cumulative return than the S&P 500, ESG 1, Quality 1 and CO2 1. The CO2 1 portfolio with the lowest emissions has a higher cumulative return compared to the CO2 3 portfolio which has the highest emissions. The Quality 3 portfolio seems to track the cumulative return of CO2 3 closely. Given the very similar cumulative return pattern for Quality 1 and ESG 3, it raises the question why they follow each other so closely. The portfolios consist of approximately 30% of the same companies each year for ESG 3 and Quality 1, which may imply that the companies with the highest individual return could be common for both portfolios. This also implies that approximately 30% of the companies that are the most profitable, as defined in chapter 3, have the lowest ESG scores every year.

However, all the selected portfolios are large portfolios consisting of many companies from different sectors. The CO2 3 portfolio may consist of many companies in the energy and industry sector, and the CO2 1 portfolios may favor companies in the technology and financial sector.

Portifolio	Expected monthly average return	Annualized return	Standard deviation of portifolio return	Annualized standard deviation	Value at risk (5%)	Conditional value at risk	Mean absolute deviation	Sharpe ratio monthly	Annualized sharpe ratio
Quality 1	1,4085 %	18,2744 %	3,6223 %	12,5481 %	-7,335 %	-7,781 %	2,720 %	0,32	1,23
Quality 3	0,6476 %	8,0544 %	4,2734 %	14,8035 %	-8,541 %	-9,957 %	3,078 %	0,10	0,35
ESG 1	0,7569 %	9,4706 %	3,8554 %	13,3553 %	-7,839 %	-8,777 %	2,7787 %	0,14	0,50
ESG 3	1,4508 %	18,8687 %	3,8367 %	13,2906 %	-6,386 %	-8,150 %	2,7821 %	0,32	1,21
CO2 1	0,9049 %	11,4159 %	3,9880 %	13,8149 %	-8,355 %	-9,299 %	2,8679 %	0,17	0,62
CO2 3	0,6283 %	7,8051 %	3,8063 %	13,1853 %	-7,589 %	-8,5707 %	2,7909 %	0,10	0,38

Table 4.29 Different measures of risk for the different portfolios.

Table 4.29 shows the expected return and the different risk measures described in chapter 3.3.2. The tables show that the portfolios have a very similar monthly standard deviation and Quality 3 has the highest. In terms of expected return, we see that Quality 1 and ESG 3 have the highest expected return. In terms of value at risk and conditional value at risk, the table shows that Quality 3 is the riskiest. In terms of the 8-year Sharpe ratio, Quality 1 and ESG 3 have the highest values compared to the other portfolios and have a more attractive risk-return relationship.

Figure 4.7 and table 4.29 indicate that the Quality 1 factor and low ESG score (ESG 3) are rewarded in the stock market. This is also supported by Franzén (2019) which finds that portfolios of low ESG score outperform portfolios of high ESG scores. The table and the figure do not contain any explicit information regarding strategies about excess returns and do not claim that these approaches will yield any excess return. The approach can rather be seen as a descriptive approach. Novy-Marx (2014) finds that the quality anomaly and gross profitability, preform relatively better than using quality strategies, such as ROA, ROE and ROIC. He also finds that the effect is strong for large-cap U.S stocks. Hirshleifer et al. (2004) argue that this is caused by investors putting too much weight on accounting profitability compared to cash profitability when forecasting future earnings, leading to biased forecasts and current market prices. Dhingra and Olson (2019) had similar findings when constructing a high and low ESG-portfolios for the S&P 500 using data from 2008-2018. Expected annual return for the portfolios were 8,23% and 8,32% with standard deviations of return of 14,91% and 16,38%, which implies that the low ESG portfolio is the riskier one for almost the same expected return.

Verheyden, Eccles and Feiner (2016) argue that fundamental information about corporate financial performance, technical information about past performance and ESG information affect stock prices. They also argue that managing carbon emissions in response to growing regulatory and social pressure arising from the threat of climate change, have an impact. This may be the reason for the low performance of the CO2 portfolios, or there might be other external factors that cause this. For example, the CO2 3 portfolio could consist of industry and energy companies and might be sensitive to the oil price, at least for some companies. A review study by Freide, Busch and Bassen (2015) of 2000 ESG studies, showed that 80% of the studies provided evidence of a positive association of various ESG measures with stock price performance. These findings contradict our findings of low (bottom 33%) ESG score yielding a high cumulative return for large European and US companies and show that the relevance of ESG is mixed.

4.8 Implications and analysis of results

The descriptive statistics for E, S, G, C and ESG score show that the average score for the sectors are very similar, but the one-way ANOVA analysis and the Granger-causality tests show that there are significant differences between some sectors. The SEM-models imply that the selected indicators are mostly reliable, but it has issues with the structural model and validity. The fixed effects panel data regression implies that ESG may be a relevant explanatory variable, given the control variables, but is dependent on model specification. The portfolio analysis shows that the Quality 1 and ESG 3 factor have the best performance in the stock market in terms of cumulative return.

These findings raise the question of how they can be related to the efficient market hypothesis, adaptative market hypothesis and use of ESG information. A strong-form hypothesis would believe that all ESG and financial information is taken into consideration and already priced in, but it is a very strict assumption. Lo (2004) introduced the Adaptive Markets hypothesis which assumes that the price reflects this information as market participants learn. If investors are successful with their ESG strategy, they are likely to try it again. If investors fail, they might try a different approach. Another outcome might be ignoring this information or abandoning it completely.

As mentioned earlier, investors might have a problem verifying ESG information. Schoenmaker and Schramade (2019) argue that ESG scores are add-ons that do not address the core issue and that they are only based on standardized reported data and policies. They also argue that the ESG scores might be biased towards large companies and they do not spot material weakness. Furthermore, they argue that based on the adaptive market hypothesis the pricing of ESG information is dependent on the number of market participants that take ESG seriously. Given bounded rationality and information asymmetry it might be hard for an individual investor to interpret this information. Another question is how one evaluates ESG information compared to fundamental financial information. Considering cognitive dissonance, one could argue that situations where tension between ESG information and how it impacts corporate financial performance could arise. One could rationalize that one or the other are the most important based on feelings or opinions. However, the magnitude of this, if any, is very hard to determine, but may affect the use of ESG information.

The fact that some variables seem to have a connection via Granger causality, and that there is a difference between this connection for different sectors, may be an indication that the ESG data are even more complex in nature. For an investor it would mean that considering the results for different companies according to certain variables is not enough, but one should also take into consideration what sector the given company is operating in. Additionally, the Granger causality effect and the added complexity it adds for certain sectors could also cloud the view of what should be evaluated in an investing decision.

A company that has been exposed to ESG controversies in the period before may be inclined to communicate in a way that it is taking steps to improve the area of controversy. The fact that there seems to be a relation between being exposed to ESG controversies and CSR strategy score the period after, can be interpreted in the direction that this is some sort of damage control or perhaps "greenwashing" by the company, to contain the damage sustained in the last period. The reason for this possible interpretation, can be seen from the variable list. CSR strategy score is defined as "the practices a company employs to communicate the integration of ESG in its day-to-day decision-making process". In other words, this may not show the actual integration of ESG, but the communication of it. Given this context, an investor should consider being wary of companies that exhibit good CSR strategy scores in a period after a bad ESG controversies score. The reason is the risk that a company merely communicates good ESG integration without having it realized. However, the Granger causality test specification has low test power and we must interpret this with caution.

One could also question if the SEM-models do measure the hypothesized latent constructs. The latent construct direct environmental exposure/impact is based on resource use and pollution. Some variables linked to pollution have been omitted due to them being too strongly related and having caused too many loadings between factors and correlated measurement errors, indicating that they all should load on one latent construct, hence the simplification of the ones directly linked to core operations. For the social and governance latent constructs we have very few indicators due to data and the companies being "too homogenous", when it might be a result of what they choose to report. This leads to the question whether the models might be a product of mixed data quality and what information that is available, combined with a sector and small sample bias.

The SEM-models imply that the latent construct for growth, based on Hamann et al. (2013), is also good indicators for companies in the S&P 500 and Stoxx 600 indices. We have used other indications of profitability compared to Hamann (2013). The models imply that ROE, ROA and ROIC are somewhat reliable indicators for profitability in the small samples, but not the largest ones. Model CFP-ESG 4 seems to be the most reliable model and implies that profitability has a weak/moderate significant negative influence on growth. However, this is a cross sectional model for 2018 and does not contain any information for different years. Even though we have a mostly reliable measurement model it does not imply that it is valid. The nonsignificant results could also be a result of term complexity.

The panel data regression results point to an overall negative connection between natural log of ESG score and stock return. Seeing this in conjunction with figure 4.7 makes the reason clearer. The companies with the highest return on their stocks accumulatively, could be the ones rated the lowest over the time-period investigated. Based on this, it is difficult to conclude with a definitive reason for the results based on the tests we have done, but there seem to be several companies having performed well in the stock market in periods where they had low ESG scores. Another reason might be how the information is interpreted and how many of the market participants that integrates the information in the decision-making process.

Traditional barriers to ESG integration such as fear of underperformance, concerns about fiduciary duty and misalignment with timeframes exist, but the most significant obstacle is obtaining quality data on ESG exposures and standards for how to use the ESG data (Eccles, Kastrapeli & Potter, 2017). In relation to the fear of underperformance, studies show that the positive relation between ESG factors and corporate financial performance takes time to be realized (Eccles, Ioannou & Serafeim, 2014). Eccles, Kastrapeli and Potter (2017) also argue, based on the global survey among asset managers and investors, that performance evaluation time frames are not well-aligned with time frames expected for achieving outperformance from ESG. Based on this, one could argue that selected time-period for the panel data regression could be too short, and wheter a cross-sectional approach using SEM-models can operationalize the relation with the lack of quality data, which we also have encountered, and ignoring development over time.

5 Conclusion

5.1 Summary and conclusion

By analyzing the relationship between corporate financial performance and environmental, social, and governance factors using aggregated and disaggregated data for companies included in the S&P 500 and Stoxx 600 indices, we have found that the relationship is dependent on level of analysis using a quantitative methodical approach.

For the disaggregated ESG data we have conducted an explorative and confirmatory factor analysis, and structural equation models to analyze the relationship. The proxies used for the latent construct corporate financial performance are growth and profitability. Both the CFA and full SEM-models resulted in mostly reliable indicators for the latent constructs, but almost none factor loadings were significant for the structural model. This could be caused by a poorly specified model and/or a lacking theoretical foundation. We could also question the validity of the hypothesized latent constructs. However, we had problems with missing data and data quality which severely limits the sample size and what it is possible to test. Most social and governance data are dichotomous which also may affect the reliability and validity. Another explanation could be that the time frame used here is not sufficient to analyze the relationship between ESG and CFP. This could be the case if the effects of ESG only materialize after a few years.

By analyzing the relationship using panel data regression, Granger causality and a portfolio approach we also see that the relation is complex. The panel data regression implies that the ESG scores are negatively related to the stock return, and the portfolio selection shows that the quality factor and lowest ESG scores (ESG 3) are the top performers in terms of cumulative return using equal weights and data from 2010-2018. The Granger-causality tests show that there are differences between sectors, but the test has limited power due to the panel data structure and the test specification.

Based on this study, the results give some implications for an investor. Firstly, it shows that the relationship is complex and trying to operationalize it to isolate the effects of E, S and G on profitability and growth, using SEM-models, based on MSCI (2020a), is challenging.

Secondly, it shows that the quality factor has performed well in the stock market and that the market not necessarily value companies with high ESG scores. However, other ESG strategies and approaches using different markets could yield different results. Another implication is that including companies from the Stoxx 600 index leads to the same conclusion as Franzén (2019), namely that companies with low ESG scores have performed better, in terms of cumulative return, than companies with high ESG scores in the stock market. This study does not conclude that ESG does not matter in relation to CFP, but the results might be dependent on the level of analysis.

5.2 Future research

In this study of the relationship between corporate financial performance and environmental, social, and governance factors we have focused on profitability and growth as latent constructs for corporate financial performance using CFA and SEM-models. Future studies could focus on different proxies/indicators for corporate financial performance and the environmental, social, and governance factors. It could be interesting to see future studies trying to operationalize liquidity and stock market performance in relationship to ESG and CFP latent constructs. Trying to operationalize other latent CFP constructs that were identified by Hamann et al. (2013) would also be interesting.

Future studies using SEM-models should focus on operationalization of the measurement instrument. We had no tested or valid measurement instrument, and had issues trying to operationalize different latent social and governance factors due to available data. Using the Granger causality tests on different overall ESG and CFP-data for different for companies in other indices could contribute to a better understanding of the relationship.

We have focused on using data from 2018 for the cross-sectional SEM- and CFA-models. For future research it could be interesting to try using other time periods or creating time series SEM-models by e.g. using average values if data is a problem. The cross-sectional design cannot explicitly say anything about causality or generalize, so it can be interesting for future research to use a time-series SEM-model or a qualitative research design. It could also be interesting to see if the model yields the same results for other samples, but missing ESG data might be a challenge. Another approach could be testing for specific sectors or industries to avoid sector bias and it might remove disturbance in the data.

Using panel data fixed effects regression and the same portfolio selection criteria, it would be interesting for future research to investigate the relevance of ESG scores in relationship to stock return and see wheter companies with low ESG scores are the top performers, in terms of cumulative return, in different markets.

References

Ahlklo, Y., & Lind, C. (2019). E, S or G? A study of ESG score and financial performance (Unpublished master`s thesis). Sweden: KTH Industrial Engineering and Management.

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*(2), 155-173.

Aouadi, A., & Marsat, S. (2018). Do ESG Controversies Matter for Firm Value? Evidence from International Data. *Journal of Business Ethics, 151*, 1027-1047.

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2014). Low-Risk Investing without Industry Bets. *Financial Analysts Journal, 70*(4), 24-41.

Auer, B. R. (2016). Do Socially Responsible Investment Policies Add or Destroy European Stock Portfolio Value? *Journal of Business Ethics, 135*, 381-397.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 16*, 74-94.

Ball, R., & Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research, 6*(2), 159-178.

Barber, B. M., & Odean, T. (1999). The Courage of Misguided Convictions. *Financial Analysts Journal, 55*(6), 41-55.

Basu, S. (1983). The relationship between earning's yield, market value and return for NYSE common stocks: Further evidence. *Journal of Financial Economics, 12*(1), 129-156.

Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample Size Effects on Chi Square and Other Statistics Used in Evaluating Causal Models. *Journal of Marketing Research, 19*(4), 425-430.

Bebchuk, L. A., Cohen, A., & Wang, C. C. (2013). Learning and the disappearing association between governance and returns. *Journal of Financial Economics, 108*, 323-348.

Bénabou, R., & Tirole, J. (2010). Individual and Corporate Social Responsibility. *Economica, 77*(305), 1-19.

Benson, B. W., & Davidson, W. N. (2010). The Relation between Stakeholder Management, Firm Value, and CEO Compensation: A Test of Enlightened Value Maximization. *Financial Management, 39*(3), 929-963.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.

BlackRock. (2018). *Sustainable investing*. Retrieved February 11, 2020, from BlackRock web site: https://www.blackrock.com/us/individual/investment-ideas/sustainable-investing

Bodie, Z., Kane, A., & Marcus, A. (2014). *Investments* (10th ed.). McGraw-Hill Education.

Bollen, K. A. (1989). A New Incremental Fit Index for General Structural Equation Models. *Sociological Methods & Research, 17*(3), 303-316.

Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin, 107*(2), 256-259.

Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modelling . In C. R, S. du Toit, & D. Sörbom, *Structural equation models: Present and future.* Chicago: Scientific Software International.

Borgers, A., Derwall, J., Koedijk, K., & ter Horst, J. (2013). Stakeholder relations and stock returns: On errors in ivestors' expectations and learning. *Journal of Empirical Finance, 22*, 159-175.

Bowen, H. S. (1953). *Social Responsibilities of the Businessman.* New York: Harper.

Bowen, N. K., & Guo, S. (2012). *Structural Equation Modeling.* Oxford University Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long, *Testing structural equation models* (p. 136). Newbury Park, CA: Sage.

Busch, T., & Hoffmann, V. H. (2011). How Hot Is Your Bottom Line? Linking Carbon and Financial Performance. *Business & Society, 50*(2), 233-265.

Carroll, A. B. (1998). The Four Faces of Corporate Citizenship. *Business and Society Review, 100-101*(1), 1-7.

Carroll, A. B. (1999). Corporate Social Responsibility: Evolution of a Definitional Construct. *Business & Society, 38*(3), 268-295.

Chen, W.-P., Chung, H., Lee, C., & Liao, W.-L. (2007). Corporate Governance and Equity Liquidity: analysis of S&P transparency and disclosure rankings. *Corporate Governance, 15*(4), 644-660.

Cho, H.-J., & Pucik, V. (2005). Relationship between innovativeness, quality, growth, profitability, and market value. *Strategic Management Journal, 26*(6), 555-575.

Chou, C.-P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle, *Structural equation modeling: Concepts, issues, and applications* (pp. 37-55). Sage Publications.

Churet, C., & Eccles, R. G. (2014). Integrated Reporting, Quality of Management, and Financial Performance. *Journal of Applied Corporate Finance, 26*(1), 56-64.

Combs, J. G., Crook, T. R., & Shook, C. L. (2005). The Dimensionality of Organizational Performance and its Implications for Strategic Management Research. *Research Methodology in Strategy and Management, 2*, 259-286.

Core, J. E., Guay, W. R., & Rusticus, T. O. (2006). Does Weak Governance Cause Weak Stock Returns? An Examination of Firm Operating Performance and Investors' Expectations. *The Journal of Finance, 61*(2), 655-687.

Crook, T. R., Ketchen, D. J., Combs, J. G., & Todd, S. Y. (2008). Strategic Resources and Performance: A Meta-analysis. *Strategic Management Journal, 29*(11), 1141-1154.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory facor analysis. *Psychological Methods, 1*(1), 16-29.

Devinney, T. M. (2009). Is the Socially Responsible Corporation a Myth? The Good, the Bad, and the Ugly of Corporate Social Responsibility. *Academy of Management Perspectives, 23*(2), 44-56.

Dhingra, G., & Olson, C. (2019, September 3). *ESG Investing: Can You Have Your Cake and Eat It Too?* Retrieved May 5, 2020, from CFA Institute Web site: https://blogs.cfainstitute.org/investor/2019/09/03/esg-investing-can-you-have-your-cake-and-eat-it-too/

Doque-Grisales, E., & Aguilera-Caracuel, J. (2019). Environmental, Social and Governance (ESG) Scores and Financial Performance of Multilatinas: Moderating Effects of Geographic International Diversification and Financial Slack. *Journal of Business Ethics*. doi:10.1007/s10551-019-04177-w

Drange, N. N., & Nath, B. (2019). Impact of Peer Effect on ESG and Pillar scores of firms (Unpublished master`s thesis). NHH. Retrieved May 6, 2020, from https://openaccess.nhh.no/nhh-xmlui/bitstream/handle/11250/2611869/masterthesis.pdf?sequence=1&isAllowed=y

Dumitrescu, E.-I., & Hurlin, C. (2012). Testing for Granger non-causality in heterogeneous panels. *Economic Modelling, 29*(4), 1450-1460.

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

Eccles, R. G., Kastrapeli, M. D., & Potter, S. J. (2017). How to Integrate ESG into Investment Decision-Making: Results of a Global Survey of Institutional Investors. *Journal of Applied Corporate Finance, 29*(4), 125-133.

Edmans, A. (2011). Does the stock market fuly value intagibles? Employee satisfaction and equity prices. *Journal of Financial Economics, 101*, 621-640.

Ehlers, T., & Packer, F. (2017, September 17). Green Bond Finance and Certification. *BIS Quarterly Review*. Retrieved from https://ssrn.com/abstract=3042378

Elkington, J. (1999). *Cannibals with forks: the triple bottom line of 21st century business.* Oxford: Capstone.

Endrikat, J., Guenther, E., & Hoppe, H. (2014). Making sense of coflicting empirical findings: A meta-analytic review of the relationship between corporate environmental and financial performance. *European Management Journal*(32), 735-751.

Fabozzi, F. J., Ma, K. C., & Oliphant, B. J. (2008). Sin Stock Returns. *Journal of Portfolio Management, 35*(1), 82-94.

Fabrigar, L. R., Wegener, D. T., Maccallum, R. C., & Strahan, E. J. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods, 4*(3), 272-299.

Fama, E. F. (1991). Efficient Capital Markets: II. *The journal of Finance, 46*(5), 1575-1617.

Fama, E. F., & French, K. R. (2014, September). A Five-Factor Asset Pricing Model. *Fama-Miller Working Paper*. Retrieved May 4, 2020, from https://ssrn.com/abstract=2287202

Fatemi, A., Glaum, M., & Kaiser, S. (2018). ESG performance andfirm value: The moderating role ofdisclosure. *Global Finance Journal, 38*, 45-64.

Festinger, L. (1957). *A theory of cognitive dissonance.* Standford University Press.

Fornell, C., & Larcker, D. F. (1981). Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research, 18*(3), 382-388.

Franzén, F. (2019). ESG Portfolios and Stock Returns: An analysis of ESGs effect on financial performance (Unpublished master`s thesis). Lund University. Retrieved May 6, 2020, from http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=8981910&fileOId=8981922

Freeman, R. E. (1984). *Strategic Management: A Stakeholder Approach.* Boston: Pitman.

Freeman, R. E., & Reed, D. L. (1983). Stockholders and Stakeholders: A New Perspective on Corporate Governance. *California Management Review, 25*(3), 88-106.

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

Friedman, M. (1970, September 13). The Social Responsibility of Business is to Increase its Profits. *The New York Times Magazine*.

Gallego-Álvarez, I., Segura, L., & Martínez-Ferrero, J. (2015). Carbon emission reduction: the impact on the financial and operational performance of international companies. *Journal of Cleaner Production, 103*, 149-159.

Gompers, P., Ishii, J., & Metrick, A. (2003). Corporate Governance and Equity Prices. *The Quarterly Journal of Economics, 118*(1), 107-156.

Graham, B. (1973). *The Intelligent Investor* (4th ed.). New York: Harpers & Row.

Granger, C. W. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica, 37*(3), 424-438.

Green, J. (2020, January 24). Goldman to Refuse IPOs If All Directors Are White, Straight Men. *Bloomberg*. Retrieved from https://www.bloomberg.com/news/articles/2020-01-24/goldman-rule-adds-to-death-knell-of-the-all-white-male-board

Greene, W. H. (2000). *Econometric Analysis* (4th ed.). Upper Saddle River, N.J: Prentice Hall.

Gu, L., & Hackbarth, D. (2013). Governance and Equity Prices: Does Transparency Matter? *Review of Finance, 17*, 1989-2033.

Guenster, N., Bauer, R., Derwall, J., & Koedijk, K. (2011). The Economic Value of Corporate Eco-Efficiency. *European Financial Management, 17*(4), 679-704.

Halbritter, G., & Dorfleitner, G. (2015). The wages of social responsibility - where are they? A critical review of ESG investing. *Review of Financial Economics, 26*, 25-35.

Hamann, P. M., Schiemann, F., Bellora, L., & Guenther, T. W. (2013). Exploring the Dimensions of Orginizational Performance: A Construct Validity Study. *Organizational Research Methods*(16), 67-87.

Hammervold, R. (2020). *Multivariate analyser med STATA* (1st ed.). Fagbokforlaget.

Hammervold, R., & Olsson, U. H. (2012). Testing structural equation models: the impact of error variances in the data generating process. *Quality & Quantity, 46*, 1547-1570.

Hassan, O. A., & Romilly, P. (2018). Relation between corporate economic performance, environmental disclosure and greenhouse gas emissions: New insights. *Business Strategy and the Environment, 27*(7), 893-909.

Hausman, J. A., & Taylor, W. E. (1981). Panel Data and Unobservable Individual Effects. *Econometrica, 49*(6), 1377-1398.

Hirshleifer, D., Hou, K., Teoh, S. H., & Zhang, Y. (2004). Do investors overvalue firms with bloatedbalance sheets? *Journal of Accounting and Economics, 38*, 297-331.

Hoang, T.-H.-V., Przychodzen, W., Przychodzen, J., & Segbotangni, E. A. (2020). Does it pay to be green? A disaggregate analysis of U.S. firms with green patents. *Business Strategy and the Environment, 29*(3), 1331-1361.

Hoelter, J. W. (1983). The Analysis of Covariance Structures: Goodness-of-Fit Indices. *Sociological Methods & Research, 11*(3), 325-344.

Hoepner, A. G., & Schopohl, L. (2018). On the Price of Morals in Markets: An Empirical Study of the Swedish AP-Funds and the Norwegian Government Pension Fund. *Journal of Business Ethics, 151*(3), 665-692.

Hoepner, A. G., & Zeume, S. (2014). Fiduciary Duty and Sin Stocks: is vice really nice? In J. P. Hawley, A. G. Hoepner, K. L. Johnson, J. Sandberg, & E. J. Waizer, *Handbook of institutional investment and fiduciary duty* (pp. 181-206). Cambridge: Cambridge University Press.

Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics, 93*(1), 15-36.

Hu, L.-t., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle, *Structural equation modeling: Concepts, issues and applications* (pp. 76-99). Sage Publications.

Hu, L.-t., & Bentler, P. M. (1998). Fit Indices in Covariance Structure Modeling: Sensitivity to Unparameterized Model Misspecification. *Psychological Methods, 3*(4), 424-453. Retrieved from https://pdfs.semanticscholar.org/a92c/9726361d9c1d165dbf2ea781b6c48363a816.pdfRelated

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Hvidkjær, S. (2017, September). ESG investing: a litterature review. Retrieved February 17, 2020, from https://dansif.dk/wp-content/uploads/2019/01/Litterature-review-UK-Sep-2017.pdf

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal Analysis: Assumptions, Models, and Data.* Beverly Hills: Sage.

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance, 48*(1), 65-91.

Jensen, M. C. (1978). Some Anomalous Evidence Regarding Market Efficiency. *Journal of Financial Economics, 6*, 95-101.

Jensen, M. C. (2002). Value Maximation, Stakeholder Theory, and the Corporate Objective Function. *Business Ethics Quarterly, 12*(2), 235-256.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications.* SPSS.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kempf, A., & Osthoff, P. (2007). The Effect of Socially Responsible Investing on Portfolio Performance. *European Financial Management, 13*(5), 908-922.

Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate Sustainability: First Evidence on Materiality. *The Accounting Review, 91*(6), 1697-1724.

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

Lai, S., Li, H., & Li, B. (2016). Does corporate governance still affect firm performance after controlling for the distress factor? *Applied Economics, 48*(13), 1197-1209.

Landi, G., & Sciarelli, M. (2019). Towards a more ethical market: the impact of ESG rating on corporate financial performance. *Social Responsibility Journal, 15*(1), 11-27.

Lewandowski, S. (2017). Corporate Carbon and Financial Performance: The Role of Emission Reductions. *Business Strategy and the Environment, 26*(8), 1196-1211.

Lo, A. W. (2004). The Adaptive Market Hypothesis. *The Journal of Portfolio Management, 30*(5), 15-29.

Lopez, L., & Weber, S. (2017). Testing for Granger Causality in Panel Data. *The Stata Journal: Promoting communications on statistics and Stata, 17*(4), 972-984.

Love, I. (2011). Corporate Governance and Performance around the World: What We Know and What We Don`t. *The World Bank Research Observer, 26*(1), 42-70.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Berlin, Heidelberg: Springer Berling Heidelberg.

Malkiel, B. G., & Fama, E. F. (1970). Efficient Captal Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25*(2), 383-417.

Mardia, K. V., & Foster, K. (1983). Omnibus tests of multinormality based on skewness and kurtosis. *Communication in Statistics - Theory and Methods, 12*(2), 207-221.

Markowitz, H. M. (1959). *Portfolio selection: efficient diversification of investments.* New Haven: John Wiley.

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle, *Structural equation modeling: Concepts, issues, and applications* (pp. 177-198). Sage Publications.

Merton, R. C. (1987). A Simple Model of Capital Market Equilibrium with Incomplete Information. *The Journal of Finance, 42*(3), 483-510.

Morton, A. (1993). Contractarianism and Rational Choice. *Analytic Philosophy, 34*(3), 177-179.

MSCI. (2020a, April). MSCI ESG RATINGS METHODOLOGY. Retrieved May 8, 2020, from https://www.msci.com/documents/1296102/14524248/MSCI+ESG+Ratings+Methodology+-+Exec+Summary+2019.pdf/2dfcaeee-2c70-d10b-69c8-3058b14109e3?t=1571404887226

MSCI. (2020b, May). *SRI World Index.* Retrieved May 18, 2020, from MSCI Web site: https://www.msci.com/documents/10199/641712d5-6435-4b2d-9abb-84a53f6c00e4

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of Goodness-of-Fit Indices for Structural Equation Models. *Psychological Bulletin, 105*(3), 430-445.

Muthén, L. K., & Muthén, B. O. (2002). How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(4), 599-620.

NBIM. (2015, December 2). *The Quality Factor (Discussion note).* Retrieved March 12, 2020, from Norges Bank Investment Management publications web site: https://www.nbim.no/en/publications/discussion-notes/2015/the-quality-factor/

Nollet, J., Filis, G., & Mitrokostas, E. (2016). Corporate social responsibility and financial performance: A non-linear and disaggregated approach. *Economic Modelling, 52*(Part B), 400-407.

Norges Bank. (2019, November 8). *Klimaendringer, klimarisiko og Norges Bank*. Retrieved March 3, 2020, from https://static.norges-bank.no/contentassets/4ea90ed3c4784ab08621e286f26cd50b/tale_plansjerematsen_08_11_2019.pdf?v=11/08/2019084257&ft=.pdf

Norman, W. (2013). Business Ethics. In H. LaFollette, *The International Encyclopedia of Ethics* (pp. 652-668). Blackwell Publishing.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics, 108*(1), 1-28.

Novy-Marx, R. (2014). Quality Investing (Working paper). Retrieved March 16, 2020, from http://rnm.simon.rochester.edu/research/QDoVI.pdf

Olsen, R. A. (1998). Behavioral Finance ans Its Implications for Stock-Price Volatility. *Financial Analysts Journal, 54*(2), 10-18.

Piotroski, J. D. (2000). Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers. *Journal of Accounting Research, 38*, 1-41.

Porter, M. E., Serafeim, G., & Kramer, M. (2019, October 16). Where ESG Fails. Retrieved April 22, 2020, from https://www.institutionalinvestor.com/article/b1hm5ghqtxj9s7/Where-ESG-Fails

Ramezani, C. A., Soenen, L., & Jung, A. (2002). Growth, Corporate Profitability, and Value Creation. *Financial Analysts Journal, 58*(6), 56-67.

Rawls, J. (1971). *A Theory of Justice.* Cambridge: Harvard University Press.

Refinitive. (2020, April). *Environmental, Social and Governance (ESG) Scores from Refinitive.* Retrieved May 4, 2020, from Refinitive Web site:

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

Riccardi, V., & Simon, H. K. (2000). What is Behavioral Finance? *Business, Education & Technology Journal, 2*(2), 1-9.

Ringdal, K. (2018). *Enhet og mangfold* (4th ed.). Fagbokforlaget.

S&P Dow Jones Indices. (2020). *Abour the S&P 500*. Retrieved February 20, 2020, from A Dow Jones Web site: https://us.spindices.com/indices/equity/sp-500

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

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chisquare statistics in covariance structure analysis. In ASA, *Proceedings of the Business and Economic Statistics Section* (pp. 308-313). Alexandria, VA: American Statistical Association.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye, & C. C. Clogg, *Latent variables analysis: Applications for developmental research* (pp. 399-419). Sage Publications.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Test of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online, 8*(2), 23-74.

Schoenmaker, D., & Schramade, W. (2019). Investing for long-term value creation. *Journal of Sustainable Finance & Investment, 9*(4), 356-377.

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling.* Mahwah, N.J: Lawrence Erlbaum Ass.

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research, 58*(7), 935-943.

Shehata, E. (2012). LMHLMXT: Stata module to Compute Breusch-Pagan Lagrange Multiplier Panel Heteroscedasticity Test. Retrieved from https://EconPapers.repec.org/RePEc:boc:bocode:s457413

Simon, H. A. (1978). Rationality as Process and as Product of Thought. *The American Economic Review, 68*(2), 1-16.

Simon, H. A. (1987). Making Management Decisions: the Role of Intuition and Emotion. *Academy of Management Perspectives, 1*(1), 57-64.

Sims, C. A. (1980). Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered. *The American Economic Review, 70*(2), 250-257.

Sloan, R. G. (1996). Do Stock Prices Fully Reflect Information in Accruals and Cash Flows About Future Earnings? *The Accounting Review, 71*(3), 289-315.

Statman, M. (1995). Behavioral Finance versus Standard Finance. *AIMR Conference Proceedings*, 14-22. doi:10.2469/cp.v1995.n7.4

Statman, M., & Glushkov, D. (2009). The Wages of Social Responsibility. *Financial Analyst Journal, 65*(4), 33-46.

Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research, 25*(2), 173-180.

Stellner, C., Klein, C., & Zwergel, B. (2015). Corporate social resposibility and Eurozone corporate bonds: The moderating role of country sustainability. *Journal of Banking & Finance, 59*, 538-549.

Stoxx. (2020, May). STOXX INDEX METHODOLOGY GUIDE (PORTFOLIO BASED INDICES). Retrieved May 14, 2020, from https://www.stoxx.com/document/Indices/Common/Indexguide/stoxx_index_guide.pdf

Studenmund, A. (2017). *A Practical Guide to Using Econometrics* (7th ed.). Pearson Education Limited.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, D.C: American Psychological Association.

Tversky, A., & Kahneman, D. (1974). Judgement under Uncertainty: Heuristics and Biases. *Science, 185*(4), 1124-1131.

UN Global Compact; International Finance Corporation; Swiss Government. (2004). *Who Cares Wins.* Retrieved January 30, 2020, from https://www.unepfi.org/fileadmin/events/2004/stocks/who_cares_wins_global_compact_2004.pdf

Utz, S. (2018). Over-investment or risk mitigation? Corporate social responsibility in Asia-Pacific, Europe, Japan, and the United States. *Review of Financial Economics, 36*(2), 167-193.

van Marrewijk, M. (2003). Concepts and Definitions of CSR and Corporate Sustainability: Between Agency and Communion. *Journal of Business Ethics, 44*, 95-105.

Velte, P. (2017). Does ESG performance have an impact on financial performance? Evidence from Germany. *Journal of Global Responsibility, 80*(2), 169-178.

Verheyden, T., Eccles, R. G., & Feiner, A. (2016). ESG for All? The Impact of ESG Screening on Return, Risk, and Diversification. *Journal of Applied Corporate Finance, 28*(2), 47-57.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle, *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Sage Publications.

Worthington, R. L., & Whittaker, T. A. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. *The counseling Psychologist, 34*(6), 806-838.

Appendix

Appendix 1: Overview of abbreviations

ESG	Environmental, social and governance factors
E	Environmental
S	Social
G	Governance
C	Controversies
CFP	Corporate financial performance
Ksi (ξ)	Latent exogenous variables
CFA	Confirmatory factor analysis
Eta (η)	Latent endogenous variables
Zeta (ζ)	Structural error associated with latent endogenous variables (error of prediction)
Beta (β)	Matrix of regression coefficients for paths between latent endogenous variables
Gamma (γ)	Matrix of regression coefficients for paths between latent exogenous variables and latent endogenous variables
Phi (ϕ)	Covariance matrix of latent exogenous variables
Psi (Ψ)	Covariance matrix of latent errors
X	Observed indicators of latent exogenous variables
Y	Observed indicators for latent endogenous variables
Delta (δ)	Measurement errors for x-indicators
Epsilon (ε)	Measurement errors for y-indicators
Lambda x (λ_x)	Matrix coefficients (factor loadings) for x-indicators
Lambda y (λ_y)	Matrix coefficients (factor loadings) for y-indicators
Theta-delta (Θ_δ)	Covariance matrix of δ (measurement errors for x-indicators)
Theta-epsilon (Θ_ε)	Covariance matrix of ε (measurement errors for y-indicators)
CR	Composite reliability
AVE	Average variance extracted
CA	Cronbach’s alpha
ROA	Return on assets

ROE	Return on equity
ROIC	Return on invested capital
Xtgcause	Panel data Granger-causality test by Lopez & Weber (2017)
SEM	Structural equation model

Appendix 2: Equations SEM-models

Equations for x and y-variables and the structual model

CFP-E 1

$x_1 = \lambda^x_{1,1}\xi_1 + \delta_1 \qquad x_2 = \lambda^x_{2,1}\xi_1 + \delta_2$

$x_3 = \lambda^x_{3,1}\xi_1 + \delta_3 \qquad x_4 = \lambda^x_{4,1}\xi_1 + \delta_4 \qquad x_5 = \lambda^x_{5,1}\xi_1 + \delta_5$

$y_1 = \lambda^y_{1,1}\eta_1 + \varepsilon_1 \qquad y_2 = \lambda^y_{2,1}\eta_1 + \varepsilon_2 \qquad y_3 = \lambda^y_{3,1}\eta_1 + \varepsilon_3$

$y_4 = \lambda^y_{4,2}\eta_2 + \varepsilon_4 \qquad y_5 = \lambda^y_{5,2}\eta_2 + \varepsilon_5 \qquad y_6 = \lambda^y_{6,2}\eta_2 + \varepsilon_6$

$\eta_1 = \gamma_{1,1}\xi_1 + \zeta_1$

$\eta_2 = \beta_{2,1}\eta_1 + \zeta_2$

CFP-E 2

$x_1 = \lambda^x_{1,1}\xi_1 + \delta_1 \qquad x_2 = \lambda^x_{2,1}\xi_1 + \delta_2$

$x_3 = \lambda^x_{3,1}\xi_1 + \delta_3 \qquad x_4 = \lambda^x_{4,1}\xi_1 + \delta_4 \qquad x_5 = \lambda^x_{5,1}\xi_1 + \delta_5$

$y_1 = \lambda^y_{1,1}\eta_1 + \varepsilon_1 \qquad y_2 = \lambda^y_{2,1}\eta_1 + \varepsilon_2 \qquad y_3 = \lambda^y_{3,1}\eta_1 + \varepsilon_3$

$\eta_1 = \gamma_{1,1}\xi_1 + \zeta_1$

CFP-E 3

$x_1 = \lambda^x_{1,1}\xi_1 + \delta_1 \qquad x_2 = \lambda^x_{2,1}\xi_1 + \delta_2$

$x_3 = \lambda^x_{3,1}\xi_1 + \delta_3 \qquad x_4 = \lambda^x_{4,1}\xi_1 + \delta_4$

$y_1 = \lambda^y_{1,1}\eta_1 + \varepsilon_1 \qquad y_2 = \eta_1$

$\eta_1 = \gamma_{1,1}\xi_1 + \zeta_1$

CFP-ESG 1

$x_1 = \lambda^x_{1,1}\xi_1 + \delta_1 \qquad x_2 = \lambda^x_{2,1}\xi_1 + \delta_2$

$x_3 = \lambda^x_{3,1}\xi_1 + \delta_3 \qquad x_4 = \lambda^x_{4,1}\xi_1 + \delta_4 \qquad x_5 = \lambda^x_{5,1}\xi_1 + \delta_5$

$x_6 = \lambda^x_{6,2}\xi_1 + \delta_6 \qquad x_7 = \lambda^x_{7,2}\xi_2 + \delta_7 \qquad x_8 = \xi_3$

$y_1 = \lambda^y_{1,1}\eta_1 + \varepsilon_1 \qquad y_2 = \lambda^y_{2,1}\eta_1 + \varepsilon_2 \qquad y_3 = \lambda^y_{3,1}\eta_1 + \varepsilon_3$

$y_4 = \lambda^y_{4,2}\eta_2 + \varepsilon_4 \qquad y_5 = \lambda^y_{5,2}\eta_2 + \varepsilon_5 \qquad y_6 = \lambda^y_{6,2}\eta_2 + \varepsilon_6$

$\eta_1 = \gamma_{1,1}\xi_1 + \gamma_{1,2}\xi_1 + \gamma_{1,3}\xi_1 + \zeta_1$

$\eta_2 = \gamma_{2,1}\xi_1 + \gamma_{2,2}\xi_1 + \gamma_{2,3}\xi_1 + \zeta_2$

CFP-ESG 2

$x_1 = \lambda^x_{1,1}\xi_1 + \delta_1 \qquad x_2 = \lambda^x_{2,1}\xi_1 + \delta_2$

$x_3 = \lambda^x_{3,1}\xi_1 + \delta_3 \qquad x_4 = \lambda^x_{4,1}\xi_1 + \delta_4 \qquad x_5 = \lambda^x_{5,1}\xi_1 + \delta_5$

$x_6 = \lambda^x_{6,2}\xi_1 + \delta_6 \qquad x_7 = \lambda^x_{7,2}\xi_2 + \delta_7 \qquad x_8 = \lambda^x_{8,3}\xi_3 + \delta_8$

$x_9 = \lambda^x_{9,3}\xi_3 + \delta_9$

$y_1 = \lambda^y_{1,1}\eta_1 + \varepsilon_1 \qquad y_2 = \lambda^y_{2,1}\eta_1 + \varepsilon_2 \qquad y_3 = \lambda^y_{3,1}\eta_1 + \varepsilon_3$

$y_4 = \lambda^y_{4,2}\eta_2 + \varepsilon_4 \qquad y_5 = \lambda^y_{5,2}\eta_2 + \varepsilon_5 \qquad y_6 = \lambda^y_{6,2}\eta_2 + \varepsilon_6$

$\eta_1 = \gamma_{1,1}\xi_1 + \gamma_{1,2}\xi_1 + \gamma_{1,3}\xi_1 + \zeta_1$

$\eta_2 = \gamma_{2,1}\xi_1 + \gamma_{2,2}\xi_1 + \gamma_{2,3}\xi_1 + \zeta_2$

CFP-ESG 3

$x_1 = \lambda^x_{1,1}\xi_1 + \delta_1 \qquad x_2 = \lambda^x_{2,1}\xi_1 + \delta_2$

$x_3 = \lambda^x_{3,1}\xi_1 + \delta_3 \qquad x_4 = \lambda^x_{4,1}\xi_1 + \delta_4 \qquad x_5 = \lambda^x_{5,1}\xi_1 + \delta_5$

$x_6 = \lambda^x_{6,2}\xi_2 + \delta_6 \qquad x_7 = \lambda^x_{7,2}\xi_2 + \delta_7 \qquad x_8 = \lambda^x_{8,3}\xi_3 + \delta_8$

$x_9 = \lambda^x_{9,3}\xi_3 + \delta_9 \qquad x_{10} = \lambda^x_{10,3}\xi_3 + \delta_{10} \qquad x_{11} = \lambda^x_{11,3}\xi_4 + \delta_{11}$

$x_{12} = \lambda^x_{12,3}\xi_4 + \delta_{12} \qquad x_{13} = \xi_5 \qquad x_{14} = \lambda^x_{14,3}\xi_5 + \delta_{14}$

$y_1 = \lambda^y_{1,1}\eta_1 + \varepsilon_1 \qquad y_2 = \lambda^y_{2,1}\eta_1 + \varepsilon_2 \qquad y_3 = \lambda^y_{3,1}\eta_1 + \varepsilon_3$

$y_4 = \lambda^y_{4,2}\eta_2 + \varepsilon_4 \qquad y_5 = \lambda^y_{5,2}\eta_2 + \varepsilon_5 \qquad y_6 = \lambda^y_{6,2}\eta_2 + \varepsilon_6$

$\eta_1 = \gamma_{1,1}\xi_1 + \gamma_{1,2}\xi_1 + \gamma_{1,3}\xi_1 + \gamma_{1,4}\xi_1 + \gamma_{1,5}\xi_1 + \zeta_1$

$\eta_2 = \gamma_{2,1}\xi_2 + \gamma_{2,2}\xi_2 + \gamma_{2,3}\xi_2 + \gamma_{2,4}\xi_2 + \gamma_{2,5}\xi_2 + \zeta_2$

CFP-ESG 4

$x_1 = \lambda^x_{1,1}\xi_1 + \delta_1 \qquad x_2 = \lambda^x_{2,1}\xi_1 + \delta_2$

$x_3 = \lambda^x_{3,1}\xi_1 + \delta_3 \qquad x_4 = \lambda^x_{4,1}\xi_1 + \delta_4 \qquad x_5 = \lambda^x_{5,1}\xi_1 + \delta_5$

$x_6 = \lambda^x_{6,2}\xi_2 + \delta_6 \qquad x_7 = \lambda^x_{7,2}\xi_2 + \delta_7 \qquad x_8 = \lambda^x_{8,3}\xi_3 + \delta_8$

$x_9 = \lambda^x_{9,3}\xi_3 + \delta_9 \qquad x_{10} = \lambda^x_{10,3}\xi_3 + \delta_{10}$

$y_1 = \lambda^y_{1,1}\eta_1 + \varepsilon_1 \qquad y_2 = \lambda^y_{2,1}\eta_1 + \varepsilon_2 \qquad y_3 = \lambda^y_{3,1}\eta_1 + \varepsilon_3$

$y_4 = \lambda^y_{4,2}\eta_2 + \varepsilon_4 \qquad y_5 = \lambda^y_{5,2}\eta_2 + \varepsilon_5 \qquad y_6 = \lambda^y_{6,2}\eta_2 + \varepsilon_6$

$\eta_1 = \gamma_{1,1}\xi_1 + \gamma_{1,2}\xi_1 + \gamma_{1,3}\xi_1 + \zeta_1$

$\eta_2 = \gamma_{2,1}\xi_2 + \gamma_{2,2}\xi_2 + \gamma_{2,3}\xi_2 + \beta_{1,2}\eta_1 + \zeta_2$

Appendix 3: Statistical tests

Panel data stationarity test, Wooldridge's test for autocorrelation in panel data, Breusch Pagan Lagrange multiplier panel heteroskedasticity test and Modified Wald test for groupwise heteroskedasticity in panel data.

Derivation test for stationarity for variables used for Granger-causality tests and test results

1) Perform ADF regressions and generate orthogalized residuals

$$\Delta_{y_{it}} = \delta_i y_{it-1} + \sum_{L=1}^{P_i} \theta_{iL} \Delta_{y_{it-L} + \alpha_{mi} d_t + \varepsilon_{it}}, \quad \text{m=1,2,3.}$$

The lag order p_i is permitted to vary across individuals.

2) $\hat{e}_{it} = \Delta y_{it} - \sum_{L=1}^{P_i} \tilde{\pi}_{iL} \Delta_{y_{it-L} - \tilde{\alpha}_{mi} d_{mt}}$

and

$$\hat{u}_{it-1} = y_{it-1} - \sum_{L=1}^{P_i} \tilde{\pi}_{iL} \Delta_{y_{it-L} - \tilde{\alpha}_{mi} d_{mt}}$$

To control for heterogeneity across individuals, e_{it} is normalized and $\hat{u}_{it-1}$ by the regression standard error from 1)

3) $\tilde{e}_{it} = \dfrac{\hat{e}_{it}}{\hat{\sigma}_{\varepsilon i}}, \quad \tilde{u}_{it-1} = \dfrac{\hat{u}_{it-1}}{\hat{\sigma}_{\varepsilon i}}$

Where $\hat{\sigma}_{\varepsilon i}$ is the regression standard error in 1).

4) $\hat{\sigma}^2_{\varepsilon i} = \dfrac{1}{T - p_i - 1} \sum_{t=p_i+2}^{T} (\hat{e}_{it} - \hat{\delta}_i \hat{u}_{it-1})^2$

5) Estimation of the long-run to short-run standard deviations

$$\hat{\sigma}_{yi}^{2} = \frac{1}{T-1}\sum_{t=2}^{T}\Delta y_{it}^{2} + 2\sum_{L=1}^{\bar{K}} w_{\bar{K}L}\left[\frac{1}{T-1}\sum_{T=2+L}^{T}\Delta_{y_{it}}\Delta_{y_{it}-L}\right]$$

6) $$s_i = \sigma_{yi} / \sigma_{\varepsilon i}$$

For each individual i, the ratio of the long-run standard deviation to the innovation standard deviation.

Panel test statistics

7) $$\tilde{e}_{it} = \delta \tilde{u}_{it-1} + \varepsilon_{it}$$

Based on a total of $N\tilde{T}$ observations

8) $$t_{\delta} = \frac{\hat{\delta}}{STD(\hat{\delta})}$$

The conventional regression t-statistic for testing δ =0 where

9) $$\hat{\delta} = \frac{\sum_{i=1}^{N}\sum_{t=2+p_i}^{T}\tilde{u}_{it-1}\tilde{e}_{it}}{\sum_{i=1}^{N}\sum_{t=2+p_i}^{T}\tilde{u}_{it-1}^{2}}$$

10) $$\mathrm{STD}(\hat{\delta}) = \hat{\sigma}_{\tilde{\varepsilon}}\left[\sum_{i=1}^{N}\sum_{t=2+p_i}^{T}\tilde{u}_{it-1}^{2}\right]^{-1/2}$$

11) $$\hat{\sigma}_{\tilde{\varepsilon}}^{2} = \left[\frac{1}{N\tilde{T}}\sum_{i=1}^{N}\sum_{t=2+p_i}^{T}(\tilde{e}_{it} - \hat{\delta}\tilde{u}_{it-1})^{2}\right]$$

Which leads to the following test-statistic:

12) $$t_{\delta}^{*} = \frac{t_{\delta} - N\tilde{T}\hat{S}_N\hat{\sigma}_{\tilde{\varepsilon}}^{-2}STD(\hat{\delta})\mu_{m\tilde{T}}^{*}}{\sigma_{m\tilde{T}}^{*}}$$

Where the mean adjustment $\mu^{*}_{m\tilde{T}}$ and standard deviation adjustment $\mu^{*}_{m\tilde{T}}$ can be given for a deterministic specification (m=1,2,3) and time series dimension $\tilde{T}$.

Test results:

H_0 : Panels contain unit root
H_1 : Panels are stationary

Variable	t-value	p-value
LNWaterWithdrawal	-34,7734	0,0000**
LNWasteTotal	-21,8324	0,0000**
LNCO2Direct	-24,7454	0,0000**
LNEnergyUse	-87,5188	0,0000**
ProductResponsibilityScore	-160,0000	0,0000**
HumanRightsScore	-64,1546	0,0000**
CSRStrategyScore	-140,0000	0,0000**
ESGScore	-31,5691	0,0000**
ESGControversiesScore	-110,0000	0,0000**

**) Significant on a 5% level

Wooldridge's test for autocorrelation in panel data:

Wooldridge's method uses the residuals from a regression in first differences. The first differencing removes the individual-level effect, the term based on the time-invariant covariates and the constant,

$$y_{it} - y_{it-1} = (X_{it} - X_{it-1})\beta_1 + \varepsilon_{it} - \varepsilon_{it-1}$$
$$\Delta y_{it} = \Delta X_{it}\beta_1 + \Delta\varepsilon_{it}$$

where Δ is the first-difference operator.

Wooldridge's procedure begins by estimating the parameters β_1 by regressing Δy_{it} on ΔX_{it} and obtaining the residuals $\hat{\varepsilon}_{it}$. If $\hat{\varepsilon}_{it}$ are not serially correlated , then $Corr(\Delta\varepsilon_{it}, \varepsilon_{it-1}) = -0,5$. Given this observation, the procedure regresses the residuals $\hat{\varepsilon}_{it}$ from the regression with first-differenced variables on their lags and tests that the coefficient on the lagged residuals is equal to -0,5. To account for the within-panel correlation in the regression of $\hat{\varepsilon}_{it}$ on $\hat{\varepsilon}_{it-1}$, the VCE is adjusted for clustering at the panel level. This test is also robust to conditional heteroskedasticity.

Modified Wald test for groupwise heteroskedasticity in a panel fixed effects model

The error process may be homoscedastic within cross-sectional units, but its variance may differ across units: a condition as groupwise heteroskedasticity. The test uses a modified Wald statistic for groupwise heteroskedasticity in the residuals of a fixed-effect regression model.

Where N_g is the number of cross sectional units. $\hat{\sigma}_i^2 = T_i^{-1} \sum_{t=1}^{T_i} e_{it}^2$ is the estimator of the ith cross-section unit's error variance, based upon the T_i residuals e_{it} available for that unit. Then define

$$\mathrm{V}_i = T_i^{-1}(T_i - 1)^{-1} \sum_{t=1}^{T_i} (e_{it}^2 - \hat{\sigma}_i^2)^2$$

As the estimated variance of $\hat{\sigma}_i^2$. The modified Wald test statistic defined as

$$\mathrm{W} = \sum_{i=1}^{N_g} \frac{(\hat{\sigma}_i^2 - \hat{\sigma}^2)^2}{V_i}$$

And distributed as $\chi^2(N_g)$ under the null hypothesis.

H_0 : Homoskedasticity
H_1 : Heteroskedasticity

Breusch Pagan Lagrange multiplier panel heteroskedasticity test

LMHLMXT: Stata module to compute Breusch-Pagan Lagrange Multiplier Panel Heteroskedasticity test. See https://econpapers.repec.org/software/bocbocode/s457413.htm.

Appendix 4: Tables SEM measurement models x and y

Measurement model observed X-variables

Model	Indicator	Parameter	Standardized factor loading	Standard error	t-value	R2
CFP-E 1						
N=151	LN CO2Direct	λ1,1	0,9	0,13	17,391	0,81
	LN WaterWithdrawl	λ2,1	0,93	0,194	12,003	0,77
	LN WasteTotal	λ3,1	0,87	0,136	16,134	0,87
	Environmental products	λ4,1	-0,22	0,037	-2,323	0,05
	Eco Design Prodcuts	λ5,1	0,19	0,038	2,462	0,04
	Cov(EnvP, Eco Des. Prod.)	Θ-δ(5,4)	0,44	0,012	6,734	-
CFP-E 2						
N=151	LNCO2 Direct	λ1,1	0,9	0,13	17,391	0,805
	LN Waste Total	λ2,1	0,87	0,194	12,003	0,765
	LN Water Withdrawl	λ3,1	0,93	0,136	16,134	0,865
	Environmental products	λ4,1	-0,22	0,037	-2,323	0,049
	Eco Design Prodcuts	λ5,1	0,19	0,038	2,462	0,036
	Cov(EnvP, Eco Des. Prod.)	Θ-δ(5,4)	0,44	0,012	6,734	-
CFP-E 3						
N=421	LN CO2 Direct	λ1,1	0,89	0,099	25,348	0,794
	LN Water Withdrawl Total	λ2,1	0,92	0,098	25,091	0,839
	LN Waste Total	λ3,1	0,81	0,109	19,425	0,654
	Eco Design Prodcuts	λ4,1	0,04	0,022	0,91	0,002
CFP-ESG						
1 N=110	LN CO2 Direct	λ1,1	0,91	0,156	14,574	0,821
	LN Waste Total	λ2,1	0,89	0,187	12,581	0,796
	LN Water Withdrawl	λ3,1	0,94	0,151	14,834	0,877
	Environmental products	λ4,1	-0,16	0,043	-1,419	0,025
	Eco Design Prodcuts	λ5,1	0,21	0,042	3,698	0,097
	Anti competition controversies	λ6,2	0,88	0,072	5,953	0,769
	Bribery & corruption controversies	λ7,2	0,91	0,075	6,127	0,836
	Policy ESG compensation	λ8,3	Fixed to 1	0,004	115,948	Fixed 1
	Cov(EnvP, Eco Des. Prod.)	Θ-δ(5,4)	0,43	0,015	5,537	
CFP-ESG						
2 N=113	LN CO2 Direct	λ1,1	0,91	0,153	14,733	0,821
	LN Waste Total	λ2,1	0,89	0,183	12,79	0,796
	LN Water Withdrawl	λ3,1	0,94	0,148	15,107	0,88
	Environmental products	λ4,1	-0,17	0,043	-1,587	0,031
	Eco Design Prodcuts	λ5,1	0,3	0,041	3,631	0,092
	Anti competition controversies	λ6,2	0,87	0,076	5,659	0,763
	Bribery & corruption controversies	λ7,2	0,91	0,079	5,782	0,823
	Board diversity	λ8,3	0,19	1,514	1,386	0,035
	Board Specific skills	λ9,3	-0,11	3,915	-0,604	0,013
	Cov(EnvP, Eco Des. Prod.)	Θ-δ(5,4)	0,45	0,015	5,821	

CFP-ESG						
3 N=113	LN CO2 Direct	λ1,1	0,91	-	-	-
	LN Waste Total	λ2,1	0,89	-	-	-
	LN Water Withdrawl	λ3,1	0,94	-	-	-
	Environmental products	λ4,1	-0,17	-	-	-
	Eco Design Prodcuts	λ5,1	0,3	-	-	-
	Anti competition controversies	λ6,2	0,79	-	-	-
	Bribery & corruption controversies	λ7,2	1	-	-	-
	Human Rights Breaches Suppliers	λ8,3	0,13	-	-	-
	Policy Child Labor	λ9,3	0,68	-	-	-
	Policy Human Rights	λ10,3	0,7	-	-	-
	Board diversity	λ11,4	0,21	-	-	-
	Board Specific skills	λ12,4	-0,06	-	-	-
	Audit Commitee Independence	λ13,5	Fixed to 1	-	-	-
	Audit Commitee Non-Exec. Member	λ14,5	0,75	-	-	-
	Cov(EnvP, Eco Des. Prod.)	Θ-δ(5,4)	0,45	-	-	-

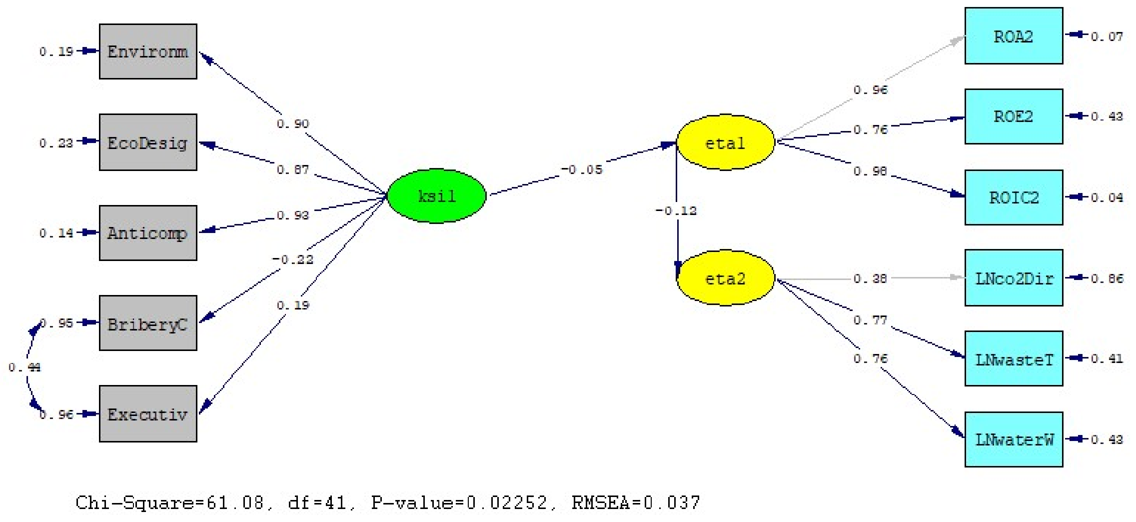
CFP-ESG						
4 N=113	LN CO2 Direct	λ1,1	0,91	0,153	14,752	0,822
	LN Waste Total	λ2,1	0,89	0,183	12,788	0,795
	LN Water Withdrawl	λ3,1	0,94	0,148	15,094	0,88
	Environmental products	λ4,1	-0,17	0,043	-1,584	0,03
	Eco Design Prodcuts	λ5,1	0,3	0,041	3,631	0,092
	Anti competition controversies	λ6,2	0,84	0,084	4,923	0,704
	Bribery & corruption controversies	λ7,2	0,94	0,474	5,094	0,892
	HumanRights Breaches Suppliers	λ8,3	0,12	0,051	1,247	0,017
	Policy Child Labor	λ9,3	0,67	0,043	3,824	0,452
	Policy Human Rights	λ10,3	0,71	0,049	2,988	0,5
	Cov(EnvP, Eco Des. Prod.)	Θ-δ(5,4)	0,45	0,015	5,819	

Measurement model observed Y-variables

Model	Indicator	Parameter	Standardized factor loading	Standard error	t-value	R2
CFP-E 1						
N=151	ROA	λ1,1	0,96	-	-	0,929
	ROE	λ2,1	0,76	2,199	8,408	0,573
	ROIC	λ3,1	0,98	0,704	21,882	0,957
	1-Year net sales/revenue growth	λ4,2	0,38	-	-	0,144
	1-Year total asset growth	λ5,2	0,77	5,278	2,629	0,594
	1-Year employee growth	λ6,2	0,76	5,377	2,487	0,573
CFP-E 2						
N=151	ROA	λ1,1	0,96	-	-	0,929
	ROE	λ2,1	0,76	2,2	8,405	0,573
	ROIC	λ3,1	0,98	0,708	21,764	0,957
CFP-E 3						
N=421	ROA	λ1,1	0,21	58,096	1,736	0,043
	ROE	λ2,1	Fixed to 1	-	-	-
CFP-ESG						
1 N=110	ROA	λ1,1	0,98	-	-	0,964
	ROE	λ2,1	0,79	1,035	21,214	0,621
	ROIC	λ3,1	0,96	0,606	25,945	0,923
	1-Year net sales/revenue growth	λ4,2	0,44	-	-	0,19
	1-Year total asset growth	λ5,2	0,72	5,147	2,695	0,522
	1-Year employee growth	λ6,2	0,89	4,802	2,843	0,787
CFP-ESG						
2 N=113	ROA	λ1,1	0,98	-	-	0,969
	ROE	λ2,1	0,79	1,044	20,189	0,623
	ROIC	λ3,1	0,96	0,561	27,763	0,92
	1-Year net sales/revenue growth	λ4,2	0,45	-	-	0,201
	1-Year total asset growth	λ5,2	0,75	5,393	2,652	0,561
	1-Year employee growth	λ6,2	0,85	4,697	2,732	0,714
CFP-ESG						
3 N=113	ROA	λ1,1	0,98	-	-	-
	ROE	λ2,1	0,79	-	-	-
	ROIC	λ3,1	0,96	-	-	-
	1-Year net sales/revenue growth	λ4,2	0,41	-	-	-
	1-Year total asset growth	λ5,2	0,68	-	-	-
	1-Year employee growth	λ6,2	0,94	-	-	-
CFP-ESG						
4 N=113	ROA	λ1,1	0,98	-	-	0,969
	ROE	λ2,1	0,79	1,006	20,89	0,622
	ROIC	λ3,1	0,96	0,537	28,935	0,919
	1-Year net sales/revenue growth	λ4,2	0,42	-	-	0,172
	1-Year total asset growth	λ5,2	0,69	4,84	2,716	0,472
	1-Year employee growth	λ6,2	0,92	5,105	2,757	0,853

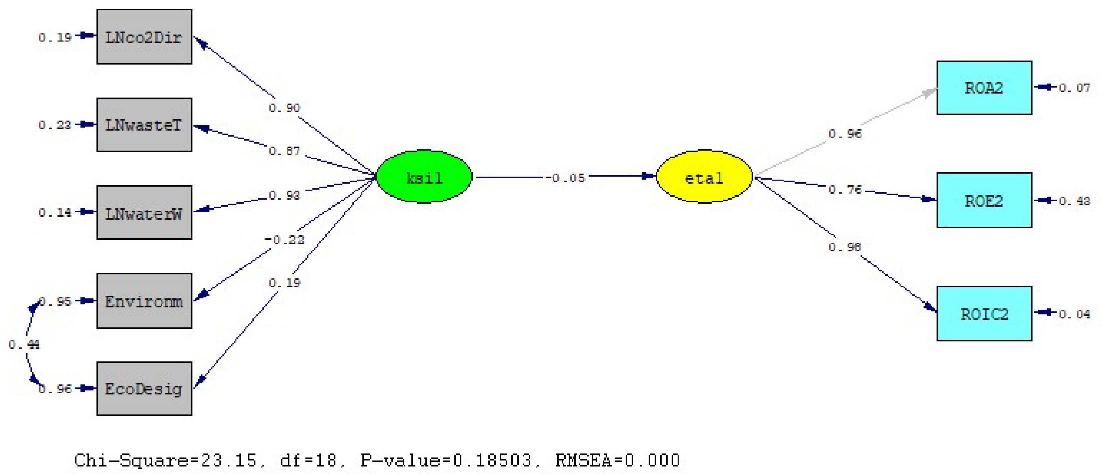
Appendix 5: Path diagrams SEM-models

CFP-E 1

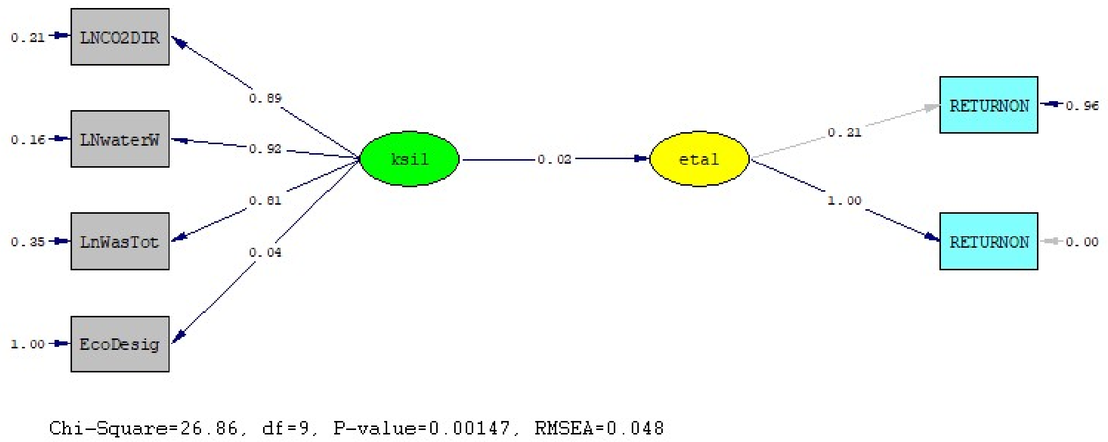


BriberyC= Eco design products and Executiv= Environmental products. The labels show wrong variable due to a LISREL program error.

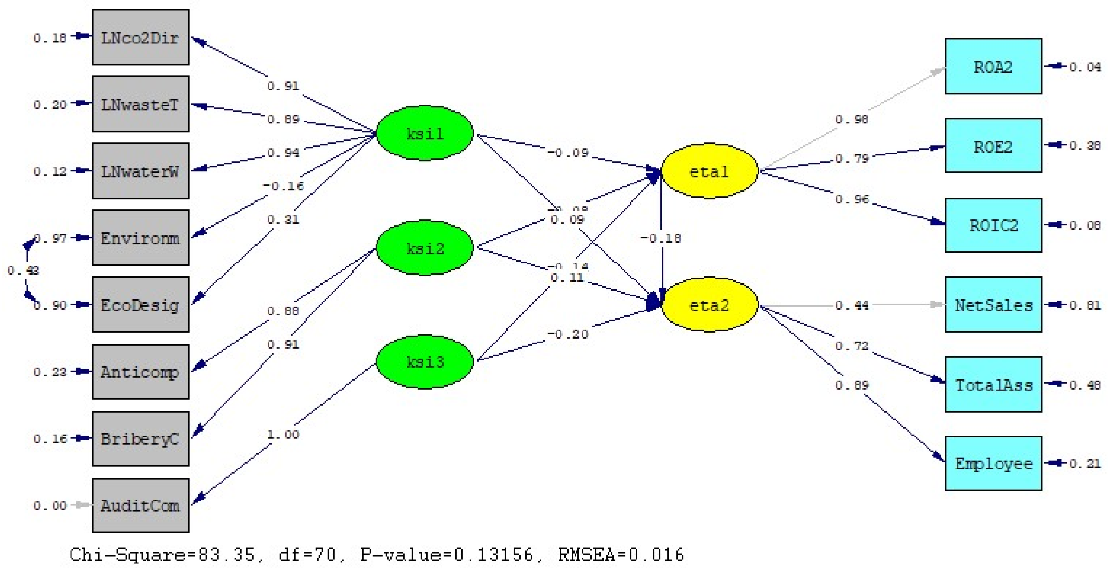
CFP-E 2



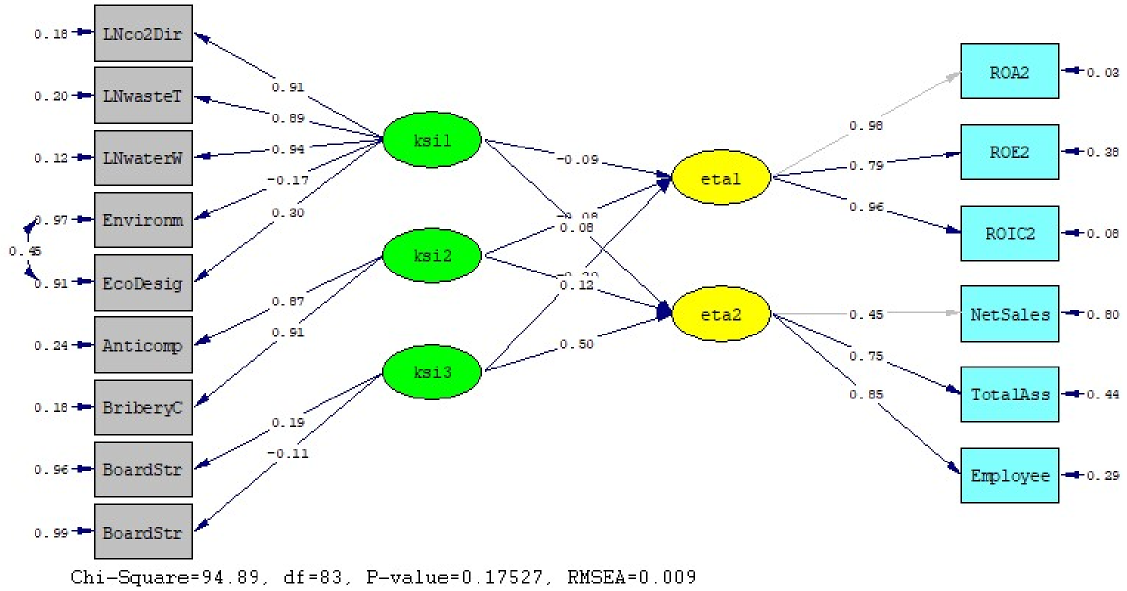
CFP-E 3



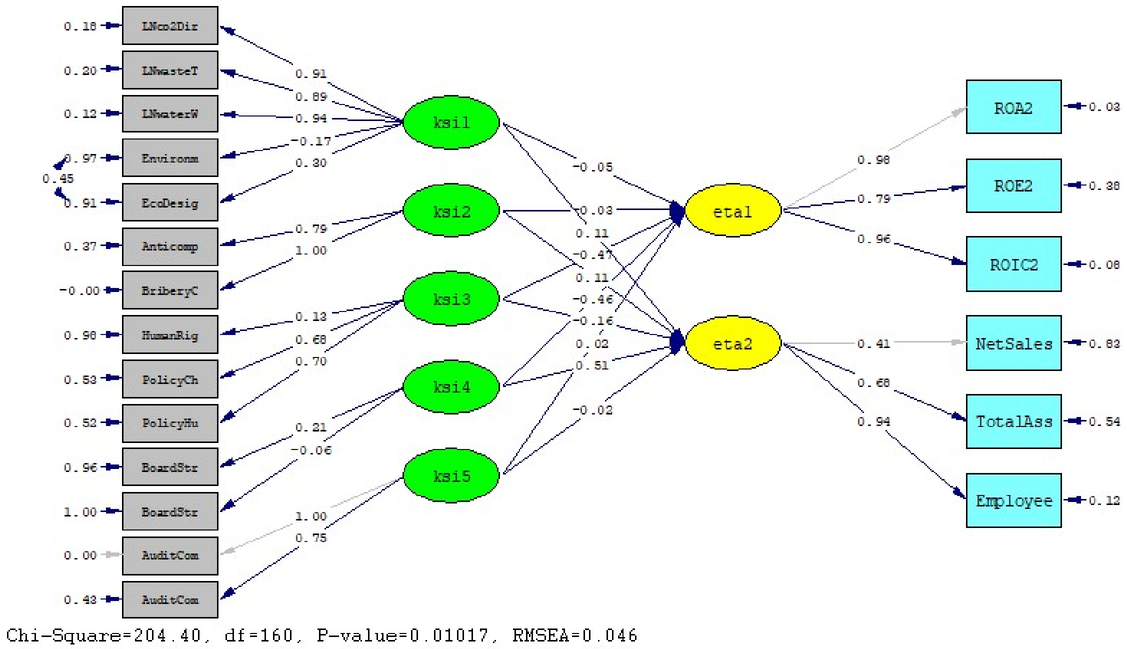
CFP-ESG 1



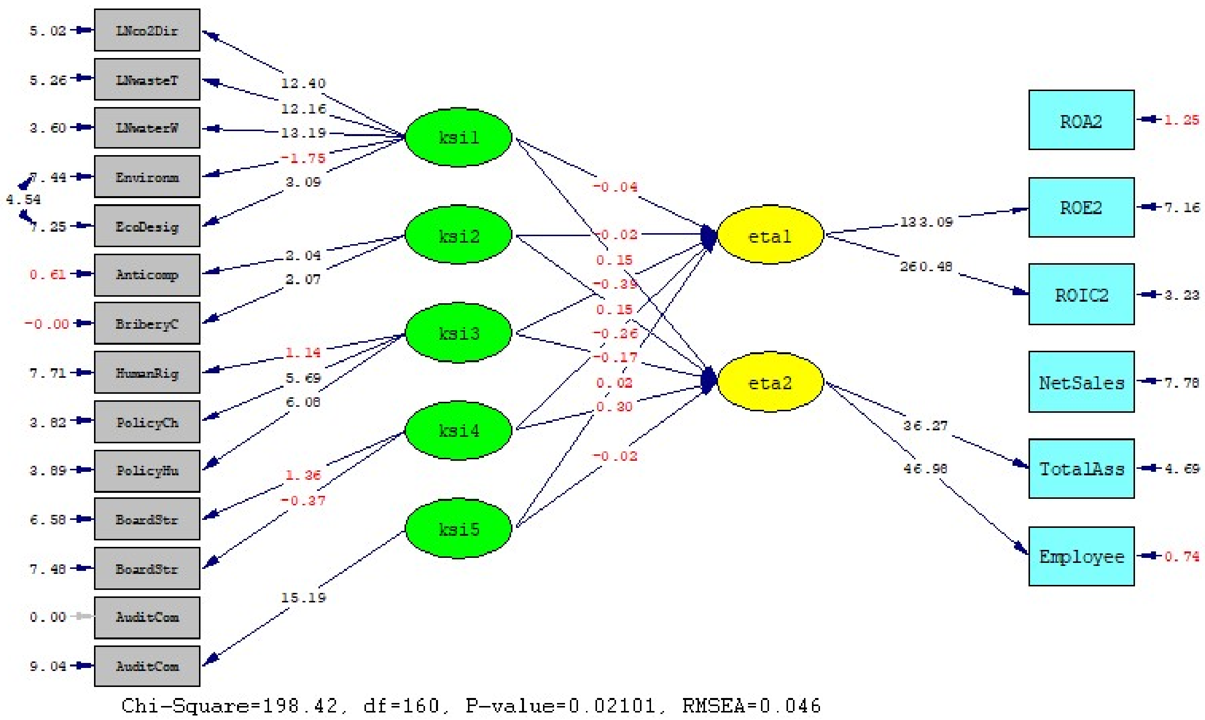
CFP-ESG 2



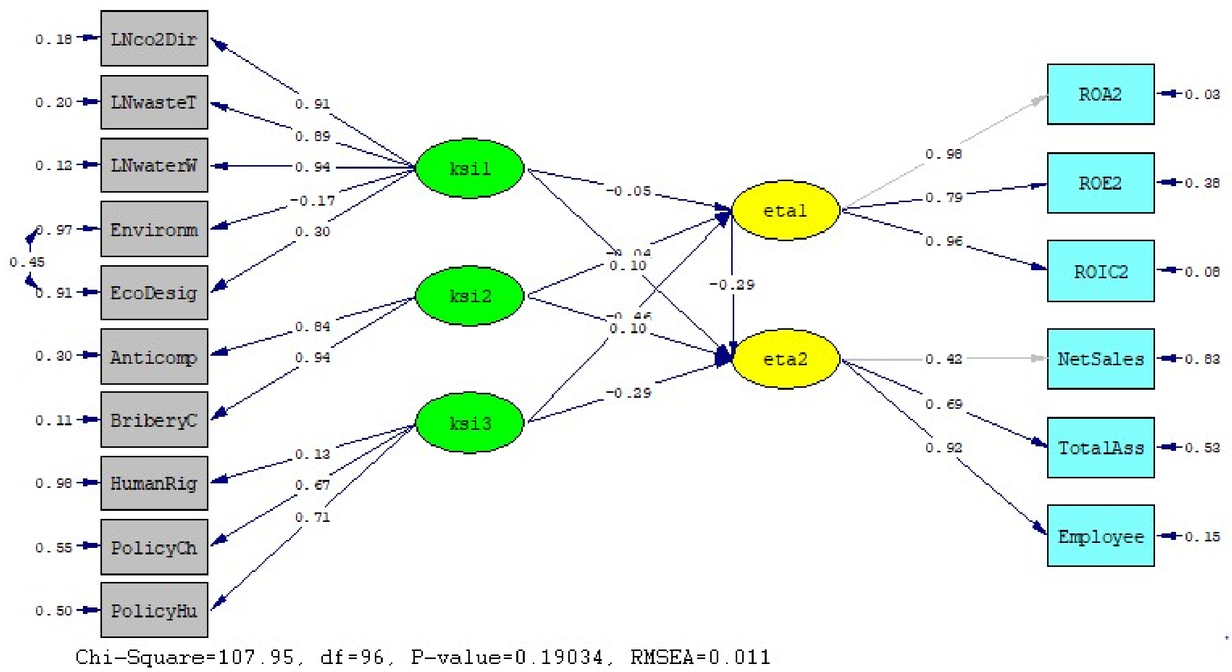
CFP-ESG 3



T-values CFP-ESG 3 non-robust estimation:


LNco2Dir
LNwasteT
LNwaterW
Environm
EcoDesig
Anticomp
BriberyC
HumanRig
PolicyCh
PolicyHu
BoardStr
BoardStr
AuditCom
AuditCom
5.02
5.26
3.60
7.44
4.54
7.25
0.61
-0.00
7.71
3.82
3.89
6.58
7.48
0.00
9.04
12.40
12.16
13.19
-1.75
3.09
2.04
2.07
1.14
5.69
6.08
1.36
-0.37
15.19
ksi1
ksi2
ksi3
ksi4
ksi5
-0.04
-0.03
0.15
-0.39
0.15
-0.26
-0.17
0.02
0.30
-0.02
eta1
eta2
133.09
260.48
36.27
46.98
ROA2
ROE2
ROIC2
NetSales
TotalAss
Employee
1.25
7.16
3.23
7.78
4.69
0.74
Chi-Square=198.42, df=160, P-value=0.02101, RMSEA=0.046

CFP-ESG 4


LNco2Dir
LNwasteT
LNwaterW
Environm
EcoDesig
Anticomp
BriberyC
HumanRig
PolicyCh
PolicyHu
0.18
0.20
0.12
0.97
0.45
0.91
0.30
0.11
0.98
0.55
0.50
0.91
0.89
0.94
-0.17
0.30
0.84
0.94
0.13
0.67
0.71
ksi1
ksi2
ksi3
-0.05
-0.04
0.10
-0.46
0.10
-0.29
-0.29
eta1
eta2
0.98
0.79
0.96
0.42
0.69
0.92
ROA2
ROE2
ROIC2
NetSales
TotalAss
Employee
0.03
0.38
0.08
0.83
0.53
0.15
Chi-Square=107.95, df=96, P-value=0.19034, RMSEA=0.011

Appendix 6: Measures of reliability

Cornabach's alpha:

$$\alpha = \frac{k * \overline{r}}{1 + \overline{r}(k-1)}$$

Where k is the number of indicators and $\overline{r}$ is the average correlation.

Composite reliability:

$$\rho_c = \frac{(\sum_i^r \lambda_i)^2}{(\sum_i^r \lambda_i)^2 + \sum_i^r \mathrm{var}(\delta_i)} \geq 0,6$$

Where λ_i is the standardized factor loading and δ_i is the measurement error.

Average variance extracted:

$$\rho_v = \frac{\sum_i^r \lambda_i^2}{\sum_i^r \lambda_i^2 + \sum_i^r \mathrm{var}(\delta_i)} \geq 0,5$$

Where λ_i is the standardized factor loading and δ_i is the measurement error.

Appendix 7: Explorative factor analysis

The rotated factor is rotated using the command "rotate, promax kaiser" in STATA 16 based on the assumption that factors could be correlated.

Environmental

Rotated factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Factor3	Factor4	Uniqueness
RenewableE~e	-0.0193	0.1357	0.0254	0.4872	0.7358
CleanTechn~y	0.0866	-0.1158	0.8742	-0.1607	0.2433
LNenergyUse	0.9462	-0.0123	-0.0356	0.0391	0.0975
LNenergyPu~t	0.9673	-0.0269	0.0420	0.0119	0.0769
ResourceEf~T	-0.0354	0.2950	0.5687	0.1314	0.5835
LNco2Direct	0.9560	-0.0156	-0.0077	-0.0395	0.0892
LNco2Indir~t	0.8636	0.0118	0.1336	0.1352	0.2240
LNco2Total	0.9821	-0.0284	0.0884	-0.0286	0.0543
LNwasteTotal	0.8363	0.0416	-0.1152	0.0356	0.2528
ResourceEf~P	0.0988	0.6731	-0.0250	0.1400	0.4860
LNwaterWit~l	0.8765	0.1311	-0.0462	-0.0541	0.1651
TargetsWat~y	0.1810	0.6693	0.1380	-0.1305	0.4931
GreenBuild~s	-0.2283	0.5514	0.3353	0.2025	0.5341
AnimalTest~g	-0.1034	0.4971	-0.3693	0.0012	0.5742
Environmen~s	-0.0719	-0.3659	0.2490	0.6821	0.2302
EcoDesignP~s	0.1719	0.0730	-0.2047	0.8317	0.2861

Rotated factor loadings:

Factor rotation matrix

	Factor1	Factor2	Factor3	Factor4
Factor1	0.9957	0.2420	-0.1266	0.0520
Factor2	0.0220	-0.3110	0.7885	0.7311
Factor3	0.0079	0.2471	0.6013	-0.5324
Factor4	-0.0897	0.8852	-0.0276	0.4235

Social

Rotated factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Uniqueness
TargetsDiv~t	0.0964	0.3896	-0.2321	0.1355	-0.0068	0.1936	0.5901	0.3317
ExecutiveM~i	0.2391	0.0086	-0.2939	0.2018	-0.0370	-0.4620	0.1914	0.5402
Strikes	-0.3174	0.4651	0.2133	0.2651	-0.0393	0.1884	0.0037	0.5502
~esSuppliers	0.2439	-0.0046	0.0644	0.5586	-0.0551	0.3682	-0.1647	0.3983
~tsSuppliers	0.0003	-0.1301	-0.1490	-0.0612	0.0528	0.8178	0.2979	0.3359
HumanRight~e	0.8411	0.0785	0.0405	0.0474	0.0909	0.1585	-0.0070	0.2067
PolicyChil~r	0.8385	-0.1248	0.0253	-0.0792	0.1216	-0.1250	0.0293	0.2571
PolicyHuma~s	0.7478	0.0521	0.1715	0.0271	-0.1904	-0.0929	0.0612	0.3690
Anticompet~y	-0.0205	0.9105	0.0642	-0.1393	0.0037	-0.0433	0.0371	0.1727
BriberyCor~n	0.0975	0.8909	-0.0198	-0.0667	0.0402	-0.1814	-0.1277	0.1840
Whistleblo~n	0.1297	0.0295	-0.1107	-0.0160	0.7819	0.1543	-0.2285	0.3351
PolicyFair~n	-0.0542	-0.1058	0.1757	0.5985	0.3626	-0.0968	-0.0097	0.4417
PolicyBrib~n	-0.0565	-0.0609	-0.1047	0.7420	-0.0614	-0.1490	0.0526	0.4118
PolicyFair~e	0.0333	-0.2164	0.3053	-0.1092	-0.0699	0.1369	0.6724	0.4549
PolicyData~y	-0.1345	0.0236	0.1169	0.0852	0.6230	-0.1207	0.3765	0.3657
ProductRes~n	0.0810	-0.0359	0.8142	0.1862	-0.0997	-0.0793	0.0188	0.2813
ProductRes~e	0.1435	0.1427	0.8108	-0.1515	0.0907	0.0190	0.1827	0.2594

Rotated factor loadings:

```
Factor rotation matrix
```

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Factor1	0.8791	0.4533	0.2816	0.0995	0.1263	0.0940	0.1706
Factor2	0.1354	-0.8105	0.4780	0.2857	0.2964	-0.1893	0.1289
Factor3	-0.4366	0.3367	0.6767	0.3219	0.0979	0.2028	0.0461
Factor4	0.0453	0.0668	-0.4201	0.8534	-0.1522	0.2384	0.1925
Factor5	-0.0539	-0.0518	-0.1478	-0.2688	0.3699	0.4660	0.5817
Factor6	-0.1022	0.1302	-0.1888	0.0602	0.6522	-0.6882	0.2911
Factor7	-0.0542	-0.0123	0.0187	-0.0261	-0.5489	-0.4081	0.7014

Governance

```
Rotated factor loadings (pattern matrix) and unique variances
```

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Uniqueness
Policy~dence	0.9708	-0.0401	0.0420	0.0860	0.0058	0.0089	0.0652
BoardFunc~ce	0.0601	0.0544	0.0682	0.0449	0.8952	0.0514	0.2409
BoardStruc~S	-0.2570	-0.1164	0.6263	0.3580	0.1715	0.0148	0.3719
BoardFunct~y	0.2782	0.2185	0.5239	-0.0229	0.0106	-0.2096	0.5627
CEOBoardMe~r	0.1159	0.2563	0.0385	-0.3309	-0.2675	0.2712	0.6173
Policy~ience	0.4651	-0.0530	0.3504	-0.1931	0.1901	0.2992	0.4898
PolBoardIndp	0.9708	-0.0401	0.0420	0.0860	0.0058	0.0089	0.0652
CSRSustain~g	-0.0326	0.0323	0.2279	-0.0478	-0.0219	-0.8590	0.2701
CSRStrateg~e	0.0087	-0.0392	0.4905	0.2591	-0.4311	0.2709	0.3374
CSRSustain~e	0.0656	0.0202	0.0971	0.7526	-0.2576	-0.0874	0.3423
PolicyESGR~n	0.0325	0.0333	0.0169	0.6561	0.2936	0.1599	0.4735
AuditCommi~2	-0.0195	0.8276	0.0857	0.1304	-0.1033	0.0369	0.2470
AuditCommi~b	-0.1132	0.8119	0.0757	0.0666	0.2071	-0.1078	0.3144
BoardStruc~i	0.0520	0.6654	-0.1152	-0.0392	-0.0206	-0.0149	0.5287
BoardStru~ty	-0.2022	0.0350	-0.7166	0.0924	0.0155	0.3343	0.3751
BoardStruc~s	-0.1566	0.5024	0.4016	-0.1947	-0.0092	0.1197	0.5383
PolicyBoar~y	0.1815	0.4197	-0.2927	0.4857	0.0230	0.0176	0.4387

Rotated factor loadings:

Factor rotation matrix

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Factor1	0.7881	0.5551	0.3974	0.0923	-0.4508	0.1658
Factor2	-0.4160	0.7658	-0.2445	0.3913	0.1266	-0.0211
Factor3	-0.3449	-0.0693	0.8653	0.1232	0.2491	-0.0513
Factor4	0.0343	-0.3152	-0.0472	0.8841	-0.1029	0.2760
Factor5	0.2859	-0.0052	-0.1633	0.1570	0.6101	-0.6387
Factor6	0.0630	0.0354	0.0676	-0.1300	0.5796	0.6966

CFP

Rotated factor loadings

Rotated factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Uniqueness
ROA2	0.9329	0.0852	-0.0136	-0.0155	0.0243	0.0749
ROE2	0.8832	-0.0457	-0.0632	-0.0117	-0.1135	0.2435
ROIC2	0.9511	0.0175	-0.0294	-0.0194	-0.0134	0.0870
NetSalesGr~h	0.0500	0.3710	0.6315	-0.1006	0.2341	0.3734
TotalAsset~h	-0.1296	0.0172	0.7987	0.1190	-0.1317	0.2893
EmployeeGr~h	-0.0364	-0.1376	0.8668	-0.1361	-0.1048	0.2792
CASHFLOWSA~S	-0.0070	0.5876	-0.1136	-0.1853	0.0853	0.5774
FREECASHFL~E	-0.1476	0.0361	-0.1509	-0.0488	0.8787	0.2510
OPERATINGP~N	0.0292	0.9422	0.0246	0.0753	-0.0361	0.1028
DIVIDENDYI~D	-0.2101	0.2263	-0.3146	-0.2564	-0.4620	0.4264
PER	-0.0960	0.0909	-0.1071	0.8537	-0.0252	0.3035
Returnonsa~s	0.0203	0.9202	0.0501	0.0917	-0.0688	0.1434
Stockreturn	0.0402	-0.0555	0.0181	0.8290	0.0226	0.2916

Factor rotation matrix

	Factor1	Factor2	Factor3	Factor4	Factor5
Factor1	0.8486	0.7069	-0.1008	-0.1729	0.0849
Factor2	0.0671	0.1463	0.8928	0.6124	0.2429
Factor3	-0.5189	0.6746	0.0148	-0.2284	-0.1752
Factor4	-0.0607	0.1505	-0.4335	0.7251	0.0393
Factor5	0.0506	-0.0340	0.0674	0.1311	-0.9495

Appendix 8: Description of variables

Variables used in Granger causality

LN Water Whitdrawal	Natural log of water whitdrawal, where water withdrawal is measured in cubic meters	ENRRDP054
LN CO2 Direct	Natural log of CO2 direct, where emissions are measured in metric tonnes. Direct emissions which are owned or controlled by the company	ENERDP024
LN Energy Use	Natural log of energy use total, which measures te total amount of energy consumed within the boundaries of the operations of the company	ENRRDP033
LN Waste Total	Natural log of waste total, which is measured by total amount of waste in metric tonnes	ENERDP045
Product Responsibility Score	A reflection of the capacity of a company to produce quality goods and services intigrating the health and safety of the customer	TRESGSOPRS
Human Rights Score	A reflection of the effectiveness of a company to respect the fundamental human rights conventions	TRESGSOHRS
ESG Controversies Score	Measure of how exposed a company is to environmental, social and governance controversies and negative events covered in global media	TRESGCCS
CSR Strategy Score	Measure of the practises a company employs to communicate that it intigrates the economic, social and environmental dimensions in its day-to-day decision-making process	TRESGCGVSS
ESG Score	Overall company score based on the self reported information in the environmental, social and governance pillars	TRESGS

Confirmatory factor analysis

Variable	Description	ASSET4 Code
LN Water Whitdrawal	Natural log of water whitdrawal, where water withdrawal is measured in cubic meters	ENRRDP054
LN CO2 Emissions Direct	Natural log of CO2 direct, where emissions are measured in metric tonnes. Direct emissions which are owned or controlled by the company	ENERDP024
LN Non-Hazardous Waste	Natural log of non-hazardous waste produced in tonnes	ENERDP049
LN Hazardous Waste	Natural log of hazardous waste produced in tonnes	ENERDP056
Eco design products	Does the company report on specific products which are designed for reuse, recycling or the reduction of environmental impacts?	ENPIDP069
Animal testing	Is the company directly or indirectly involved in animal testing?	ENPIDP057
Renewable/clean energy products	Does the company develop products or technologies for use in clean, renewable energy production?	ENPIDP066
Health & safety policy	Does the company have a policy to improve employee health & safety within the company and its supply chain?	SOHSD01V
Policy diversity and opportunity	Does the company have a policy to drive diversity and equal opportunity?	SODODP0081
Strikes	Has there been a strike or an industrial dispute that led to lost working days?	SOEQDP037
Policy human rights	Does the company have a policy to ensure the respect of human rights in general?	SOHRDP0105
Policy fair competition	Does the company describe in the code of conduct that it strives to be a fair competitior?	SOCODP0066
Policy bribery and corruption	Does the company describe in the code of conduct that it strives to avoid bribery and corruption in all its operations?	SOCODP0067
Whistleblower protection	Does the company have a provision or comply with regulations protecting whistleblowers?	SOCODP011
Policy cutomer health & safety	Does the company have a policy to protect cutomer health & safety?	SOPRDP0121
Policy data privacy	Does the company have a policy to protect cutomer and general public privacy and integrity?	SOPRDP0124
Policy responsible marketing	Does the company have a policy on responsible marketing ensuring protection of children?	SOPRDP0126
Policy fair trade	Does the company have a policy on fair trade?	SOPRDÅ0128
Audit commitee independence	Percentage of independent board members on the audit committee as stipulated by the company	CGBFO01V
Audit commitee non-executive	Percentage of non-executive board members on the audit committee as stipulated by the company	CGBFDP019

Compensation commitee independence	Percentage of independent board members on the compensation committee as stipulated by the company	CGBFO04V
Compensation commitee non-executive	Percentage of non-executive board members on the compensation committee as stipulated by the company	CGBFDP021
Board structure gender diversity	Percentage of female on the board	CGBSO03V
Board specific skills	Percentage of board members who have either an industry specific background or a strong financial background	CGBSO04V
Board members strictly independent	Percentage of independent board members as reported by the company	CGBSO07V
CSR sustainability commitee	Does the company have a CSR committee or team?	CGVSDP005
CSR reporting global activities	Does the extra financial report of the company take into account the global activities of the company	CGVSDP029
CSR sustainability external audit	Does the company have an external auditor of their CSR/H&S/Sustainability report?	CGVSDP030
ESG reporting scope	The percentage of the activities of the company covered in its environmental and social reporting	CGVSDP041
Annual stock return **Δ**	(Price t- price t-1) /price t-1)	P
Price/earnings ratio	The price divided by the earnings ratio per share at the given date	PE
Dividend yield	The dividend yield expresses the dividend per share as a percentage of the share price	DY
1-Year net sales/revenue growth	(Current year`s net sales or revenues / last year`s total net sales or revenues - 1) * 100	WC08631
1-Year total asset growth	(Current year`s total assets / last year`s total assets - 1) * 100	WC08621
1-Year employee growth	(Current year`s total employees / last year`s total employees - 1) * 100	WC08626

Δ Note that the ASSET4 code for this vaiable is for price, and not for annual stock return directly, as the stock return was calculated manually using the price variable

Structural equation models (SEM)

Variable	Description	ASSET4 Code
LN Water Whitdrawal	Natural log of water whitdrawal, where water withdrawal is measured in cubic meters	ENRRDP054
LN CO2 Direct	Natural log of CO2 direct, where emissions are measured in metric tonnes. Direct emissions which are owned or controlled by the company	ENERDP024
LN WasteTotal	Natural log of total amount of waste produced in tonnes	ENERDP045
Environmental products	Does the company report on at least one product line or service that is designed to have positive effects on the environment or which is environmentally labeled and marketed?	ENPIDP019
Eco design products	Does the company report on specific products which are designed for reuse, recycling or the reduction of environmental impacts?	ENPIDP069
Anti competition controversies	Is the company under the spotlight of the media because of a controversy linked to anti-competitive behaviour (e.g. anti-trust and monopoly), price-fixing or kickbacks?	ECCLO13V
Bribery and corruption controversies	Is the company under the spotlight of the media because of a controversy linked to bribery and corruption, political contributions, improper lobbying, money laundering, parallell imports or any tax fraud?	SOCOO10V
Policy ESG compensation	Does the company have an extra-financial performance oriented compensation policy?	CGCPDP0013
Board diversity	Percentage of female on the board	CGBSO03V
Board Specific skills	Percentage of board members who have either an industry specific background or a strong financial background	CGBSO04V
Human Rights Breaches Suppliers	Does the company report or show to be ready to end a partnership with a sourcing partner if human rights criteria are not met?	SOHRDP029
Policy Child Labor	Does the company have a policy to avoid the use of child labor?	SOHRDP0102
Policy Human Rights	Does the company have a policy to ensure the respect of human rights in general?	SOHRDP0105
Audit Commitee Independence	Percentage of independent board members on the audit commitee as stipulated by teh company	CGBFO01V
Audit Commitee Non-Exec. Member	Percentage of non-executive board members on the audit commitee as stipulated by the company	CGBFDP019

Return on assets (ROA)	(Net income - bottom line + ((interest expense on debt-interest capitalized) * (1 - tax rate))) / averege of last year`s and current year`s total assets * 100	WC08326
Return on equity (ROE)	(Net income - bottom line - preferred dividend requirement) / average of last year`s and current year`s common equity * 100	WC08301
Return on invested capital (ROIC)	(Net income - bottom line + ((interest expense on debt - interest capitalized) * (1 - tax rate))) / averege of last year`s and current year`s (total capital + short term debt & current portion of long term debt) * 100	WC08376
1-Year net sales/revenue growth	(Current year`s net sales or revenues / last year`s total net sales or revenues - 1) * 100	WC08631
1-Year total asset growth	(Current year`s total assets / last year`s total assets - 1) * 100	WC08621
1-Year employee growth	(Current year`s total employees / last year`s total employees - 1) * 100	WC08626

Panel data regression

Variable	Description
ESG score	An overall company score based on the self-reported information in the environmental, social and governance pillars
Natural log ESG score	Natural log of the ESG score
Natural log Lag(1) ESG score	Natural log and one time lag of the ESG score
Natural log controversies score	Natural log of the measure of a company`s exposure to to environmental, social and governance controversies and negative events reflected in global media
Natural log environmental score	Natural log of the weighted average relative rating of a company based on the reported environmental information and the resulting three environmental category scores
Natural log social score	Natural log of the weighted average relative rating of a company based on the reported social information and the resulting four social category scores
Natural log governance score	Natural lof of the weighted average relative rating of a company based on the reported governance information and the resulting three governance category scores
Return on equity (ROE)	(Net income - bottom line - preferred dividend requirement) / average of last year`s and current year`s common equity * 100

Return on assets (ROA)	(Net income - bottom line + ((interest expense on debt-interest capitalized) * (1 - tax rate))) / averege of last year`s and current year`s total assets * 100
Lag(1) ROE	The regular formula for ROE, time lagged t-1
Lag(1) ROA	The regular formula for ROA, time lagged t-1
Natural log capital expenditure/assets	Natural log of (Capital expenditures / last year`s (total assets - cutomer liabilities on acceptances) * 100)
Natural log assets	The natural logarithm of the total amount of assets in a given company
Natural log debt/assets	Natural log of (Short term debt & current portion of long term debt + long term debt) / total assets * 100)
Lag(1) Cash flow operations/sales	Time lagged (funds from operations / net sales or revenues * 100)
Natural log Lag(1) cash flow operations/sales	Natural log of (funds from operations / net sales or revenues * 100)

Appendix 9: Correlation matrix panel data fixed effects regression models

	ROE	ROA	ROIC	Simple~n	LN_CFs~s	LN_deb~s	LNCScore
ROE	1.0000						
	9417						
ROA	0.5264	1.0000					
	0.0000						
	9366	9551					
ROIC	0.6004	0.7530	1.0000				
	0.0000	0.0000					
	9382	9509	9559				
Simplereturn	0.0349	0.0864	0.0380	1.0000			
	0.0009	0.0000	0.0003				
	9056	9162	9172	9213			
LN_CFsales	0.0080	0.0348	0.0209	-0.0168	1.0000		
	0.4471	0.0009	0.0452	0.1138			
	9010	9138	9148	8834	9407		
LN_debtAss~s	0.0332	-0.0047	-0.0068	-0.0102	0.0443	1.0000	
	0.0018	0.6577	0.5178	0.3400	0.0000		
	8851	8978	8990	8689	9013	9236	
LNCScore	0.0043	0.0646	0.0337	0.0778	0.0260	0.0009	1.0000
	0.6891	0.0000	0.0016	0.0000	0.0167	0.9318	
	8664	8764	8776	8695	8451	8309	8818

LNEscore	0.0076	-0.0538	-0.0046	-0.1171	-0.0855	0.0206	-0.2317
	0.4825	0.0000	0.6653	0.0000	0.0000	0.0609	0.0000
	8641	8740	8754	8683	8429	8288	8793
LNSscore	0.0158	-0.0372	-0.0121	-0.1002	-0.0654	0.0276	-0.2728
	0.1423	0.0005	0.2579	0.0000	0.0000	0.0121	0.0000
	8641	8740	8754	8683	8429	8288	8793
LNGscore	0.0313	-0.0308	0.0069	-0.1115	-0.0200	0.0317	-0.1627
	0.0036	0.0040	0.5192	0.0000	0.0665	0.0039	0.0000
	8642	8741	8755	8684	8430	8289	8793
LNAssets	-0.0347	-0.1263	-0.0491	-0.0798	0.0262	0.0660	-0.2277
	0.0008	0.0000	0.0000	0.0000	0.0111	0.0000	0.0000
	9237	9371	9379	9059	9397	9236	8654
LN_CFsales	0.0080	0.0348	0.0209	-0.0168	1.0000	0.0443	0.0260
	0.4471	0.0009	0.0452	0.1138	0.0000	0.0000	0.0167
	9010	9138	9148	8834	9407	9013	8451
LNcapExpAs~s	0.0333	0.0891	0.0447	0.0119	-0.0322	0.1367	0.0230
	0.0021	0.0000	0.0000	0.2760	0.0026	0.0000	0.0390
	8547	8686	8683	8404	8738	8537	8022

	LNEscore	LNSscore	LNGscore	LNAssets	LN_CFs~s	LNcapE~s
LNEscore	1.0000					
	8793					
LNSscore	0.6720	1.0000				
	0.0000					
	8793	8793				
LNGscore	0.3289	0.3913	1.0000			
	0.0000	0.0000				
	8793	8793	8794			
LNAssets	0.1856	0.1988	0.1145	1.0000		
	0.0000	0.0000	0.0000			
	8631	8631	8632	9653		
LN_CFsales	-0.0855	-0.0654	-0.0200	0.0262	1.0000	
	0.0000	0.0000	0.0665	0.0111		
	8429	8429	8430	9397	9407	
LNcapExpAs~s	-0.0308	-0.0400	-0.0047	-0.2660	-0.0322	1.0000
	0.0059	0.0003	0.6723	0.0000	0.0026	
	8004	8004	8005	8905	8738	8905

```
. pwcorr Simplereturn LNesgScore LNCScore LNEscore LNSscore LNGscore, obs sig
```

	Simple~n	LNesgS~e	LNCScore	LNEscore	LNSscore	LNGscore
Simplereturn	1.0000					
	9213					
LNesgScore	-0.1364	1.0000				
	0.0000					
	8684	8794				
LNCScore	0.0778	-0.2790	1.0000			
	0.0000	0.0000				
	8695	8793	8818			
LNEscore	-0.1171	0.8492	-0.2317	1.0000		
	0.0000	0.0000	0.0000			
	8683	8793	8793	8793		
LNSscore	-0.1002	0.8638	-0.2728	0.6720	1.0000	
	0.0000	0.0000	0.0000	0.0000		
	8683	8793	8793	8793	8793	
LNGscore	-0.1115	0.6748	-0.1627	0.3289	0.3913	1.0000
	0.0000	0.0000	0.0000	0.0000	0.0000	
	8684	8794	8793	8793	8793	8794

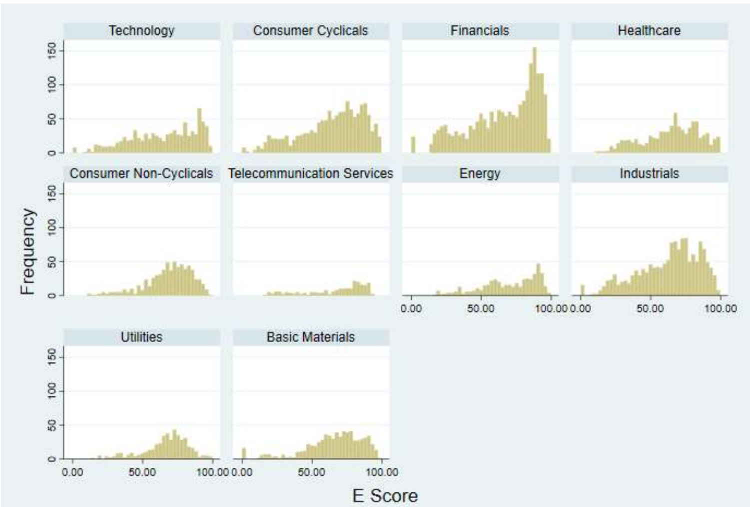
Appendix 10: Conversion monthly return and standard deviation to annualized.

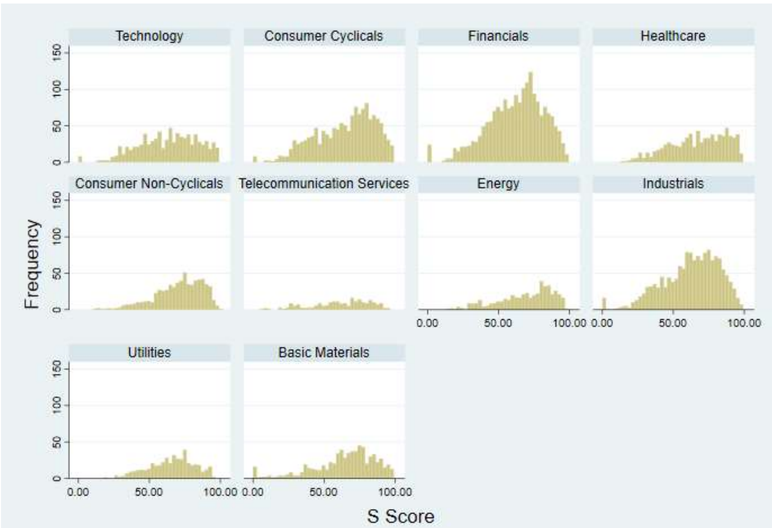
Monthly to annual return: $((1+r_M)^{12}-1)*100$

Monthly standard deviation to annual return**:** $\sqrt{12}\ *\sigma_M$

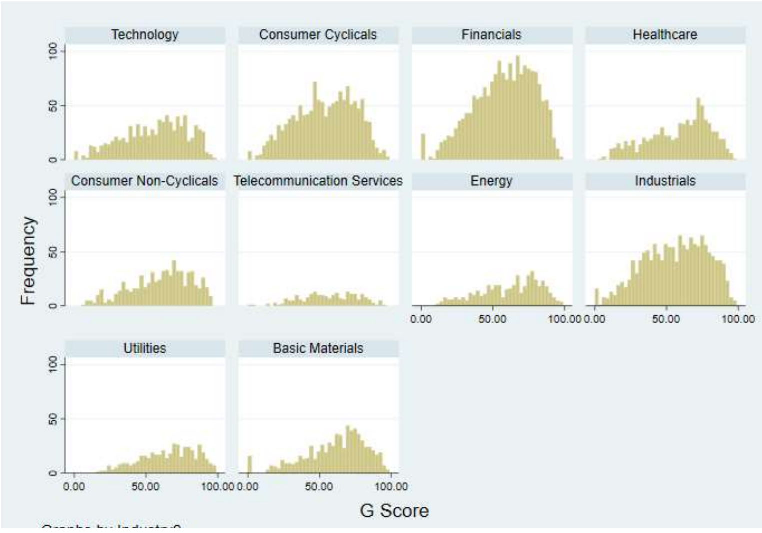
Appendix 11: Charts distribution Environmental, social, governance and controversies score

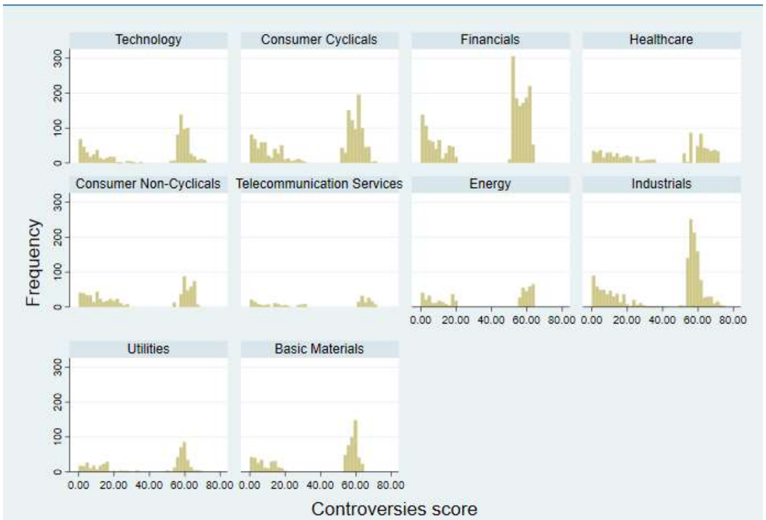
Environmental score and social score





Governance score and controversies score





Håvard Lismoen, Peder Martin Moum

The relationship between ESG and corporate financial performance

NTNU

Norwegian University of Science and Technology