Universitat de Barcelona

Facultat d'Economia i Empresa

ESG PERFORMANCE AND STOCK MARKET RETURNS

Philipp Mühr

Professor: Montserrat Termes-Rifé

Bachelor of Economics

ABSTRACT

This paper shows that stock market returns are related to ESG (Environment, Social and Governance) performance and that different areas of ESG performance affect returns in different ways. It shows that improved environmental performance is related with decreased returns, likely due to decreased idiosyncratic risk. At the same time improved social and governance performance is related to increased returns, meaning that these factors are indicators of quality. These insights are based on data from German and European stocks for the period from 2000 to 2020. Fixed-effect least square estimation was utilized. Additionally, the paper shows that there are some differences in these relationships depending on the size of firms. It also shows that these relationships are stronger in the time period after the financial crisis than for the period before the financial crisis.

Keywords: ESG (Environment, Social and Governance), CSR (corporate social responsibility), SRI (socially responsible investment), sustainable finance, stock market analysis

RESUMEN

Este trabajo muestra que los rendimientos del mercado de valores están relacionados con el desempeño de ESG (Medio Ambiente, Social y Gobernanza empresarial) y que las diferentes áreas de ESG afectan a los rendimientos de diferentes maneras. Muestra que el desempeño ambiental mejorado está relacionado con una reducción de los rendimientos, probablemente por motivo de disminución del riesgo idiosincrásico. Al mismo tiempo, el mejor desempeño social y de gobernanza está relacionado con mayores rendimientos, lo que significa que estos factores son indicadores de calidad. Estos resultados se basan en datos de acciones alemanas y europeas para el período de 2000 a 2020. Se utilizó modelos de efecto fijo. Además, el trabajo muestra que existen algunas diferencias en estas relaciones en función del tamaño de las empresas. También muestra que estas relaciones son más fuertes en el período posterior a la crisis financiera que en el período anterior a la crisis financiera.

Palabras clave: ESG (Medio Ambiente, Social y Gobernanza empresarial), CSR (Responsabilidad Social Corporativa), SRI (Inversión Socialmente Responsable), finanzas sostenibles, análisis del mercado de valores

INDEX

LIST OF FIGURES AND TABLES

I. INTRODUCTION

This paper aims to investigate if ESG (Environment, Social and Governance) ratings are related to excess returns on a given stock and if they are a valid indicator for current and future performance. Furthermore, it aims to evaluate if these effects vary depending on the size of a given firm and if these correlations have changed over time, due to the increased interest in socially responsible investment in recent years.

To some extend this investigation is only possible, because of the recent explosion in data being collected and published on this topic. An increasing number of firms have started publishing voluntary reports on their ESG performance since the early 2000s, which has further expanded within the European Union since the adoption of the non-financial reporting directive (Directive 2014/95/EU) making such publication mandatory for some firms since 2018. At the same time fundamental changes in collective investment have taken place. The proportion of capital being managed with sustainability goals in mind has increased considerably in the past decade, both through private special interest funds as well as adoption of sustainability goals by public investors (Puaschunder, 2016).

This trend is likely to continue over the coming years, in light of the implementation of further legislation in this area, such as the EU's Green Finance Initiative, which would require certain institutional investors to create and follow ESG strategies. It, therefore, stands to reason that it should be evaluated if and how the ESG performance of a given firm relates to their stock market performance, something which depends both on what ESG ratings say about a given firm and how investors utilise this information.

This paper utilizes the following structure: The chapter entitled Background, discusses the history of ESG and CSR (Corporate Social Responsibility) as a topic and how it has changed over time. It also highlights how different researchers have approached the topic. In the following chapter, Methodology, the approach taken by this paper is detailed and the variables and data used are described. In Results the empirical results of this paper are described at length. In the chapter Discussion, the results of this paper will be related to other works and potential interpretations will be discussed. In the Conclusion the results are presented in a concise manner. The Annex contains summary statistics and an overview of figures, tables and variables.

II. BACKGROUND

2.1 Overview

Over the last two decades and more strongly since the great recession, corporate ESG strategy and sustainability have become important factors in how capital is allocated. While investors choosing where to place their capital have often taken into account additional goals apart from maximizing financial returns, doing so now has much broader appeal and is done in much more sophisticated ways than before. From the turn of the 21st century onwards the number of funds and institutional investors integrating ESG criteria in their selection processes has increased rapidly, especially within the last decade (Eccles, Kastrapeli and Potter, 2017). At the same time data providers have widened their coverage to include more metrics outside of traditional financial information. Additional reporting requirements have further increased the availability of information on company conduct (Christensen, Hail and Leuz, 2019).

Research on ESG and CSR (Corporate Social Responsibility) goes back more than four decades and ultimately is an extension of the long-standing debate about what the role of corporations within society should be. One side holds the view that firms should take into account social welfare and that investors should avoid certain investments, a prominent early example of this would be bonds supporting the apartheid regime in South Africa. The opposing view holds that profit maximisation should be the sole objective of firms, as it was famously argued by Milton Friedman (Friedman, 1970).

More recently and taking the perspective of the investor side, no consensus has been reached on the question of how portfolio construction taking into account ESG criteria affects financial performance. The popular view — that arbitrary restrictions on the potential investment universe can only negatively affect potential returns — fails to consider that restrictions based on ESG criteria may not be arbitrary at all, but are likely related to how a firm operates and how it is seen by the public. Accordingly, how ESG performance relates to the return and risk of any given investment remains uncertain. Arguably this is not all that surprising, when considering that accurate modelling of investments, particularly in equity markets, still faces a number of barriers, with many competing, but imperfect, approaches being used.

Despite that, many attempts have been made at explaining how ESG performance relates to the risk and performance of an underlying investment. The large number in variables available, and the introduction of comparable, though imperfect, ESG scores, which are the results of complex data gathering processes, have made this kind of analysis much easier than it used to be even a few years ago. For investors interested in SRI (Socially Responsible Investment) strategies they also reduce information cost and enable more nuanced selection strategies going beyond simple exclusion criteria.

2.2 Review of literature and methodology

Empirical results still come to different and often times contrary results. Some of this is due to different markets being analysed, different indicators being chosen, differences in the underlying assumptions and the general difficulty in analysing financial markets. Further complicating the matter, the way in which ESG scores are constructed varies widely between data providers, including large differences in methodology. Some scores are based solely on publicly available information, others use surveys and questionnaires. They also differ in the choice and weighting of metrics within the score and how they impute and/or penalize missing data. Almost all scores use a best in class approach, where scores are relative in nature with respect to a given firms peer group. How these peer groups are defined can also vary considerably (Kotsantonis and Serafeim, 2019). This means that scores by different providers are difficult to compare directly and may have different blind spots. In practice using ESG scores is difficult to avoid, as there are few other metrics available or the metrics, that are available, are too limited in scope. For the environmental impact of a firm CO_2 equivalent emissions are often the primary choice, but it is difficult to decide what the scope of emissions to be included should be. This could range from just direct emissions, to including any emissions generated throughout the whole product lifecycle, emissions generated by suppliers and emissions of activities financed by a company. It also ignores other types of pollution such as soil and water pollution. The same problem applies even more to social- and governance factors, with many potential datapoints being confidential. The few publicly available statistics, such as workplace deaths, are hardly comparable between industries.

Within the literature, there are three primary methodological approaches used for this kind of analysis. One is using regression models of varying complexity to estimate coefficients that describe the relationships between variables. There are some obstacles to this, particularly regarding the choice of control variables and the distributional properties of financial returns. Another popular approach is creating a number of portfolios and comparing their relative performance, generally as an extension to the Fama-French three- or five-factor models (Fama and French, 1993, 2014). While this approach is an efficient way to reduce volatility and, therefore, get more stable results, there is some uncertainty about how the resulting coefficients are to be interpreted. This type of model is also dependant on how exactly portfolios are constructed and how the market is defined. The third approach is to use specific events as natural experiments and to see if firms fare differently after such an occurrence. Most often, changes in policy, such as new publication requirements or expenditure obligations are used (Manchiraju and Rajgopal, 2017). The problem with this approach is that, in practice, such studies always estimate the effect of whichever policy was applied and not the effect of ESG performance in and of itself. As a result, it is impossible to use this approach to gauge how ESG performance and investor preference interact, and how this relation may have gradually changed over time. For the purposes of this paper a regression analysis approach is chosen, with some measures to resolve the potential issues that have been described.

2.3 Distinguishing between CSR and ESG

The meaning of CSR (Corporate Social Responsibility) and ESG (Environment, Social and Governance) has changed over time with much more focus being placed on environmental and climate concerns recently. In practice, both terms are often used interchangeably. For the purposes of this document a distinction shall be made to provide more clarity. CSR is the older term used since the 1950s, while the acronym ESG was coined in 2005 as a result of an UN initiative, which has resulted in more precisely defined objectives and criteria for the latter. CSR often entails firms contributing to philanthropic goals, applying their workforce to volunteer work or supporting charities and similar organisations (Garriga and Melé, 2004). ESG performance, as the term is used currently, focuses on reducing (environmental) externalities, on treatment of the workforce and supply chain and on accurately following governance and reporting standards (van Duuren, Plantinga and Scholtens, 2016). The term CSR often is considered to include most of the areas named under ESG performance, but the reverse is usually not the case.

With regard to classic CSR activities views in the literature diverge to a notable extend. Some authors argue that actively following CSR is not consistent with the primary goal of maximising profits and therefore return for shareholders, falling in line with Milton Friedman's view. In this view, CSR activities could be the result of agency problems, where they benefit management rather than shareholders; or these activities simply are distractions that reduce managerial performance and waste company resources (Benabou and Tirole, 2010; Masulis and Reza, 2015; Xu and Genton, 2015). The alternative view holds that CSR activities are indicators of good governance and could positively affect value by improving stakeholder relations and brand image (Dickinson, 1990; Ferrell, Liang and Renneboog, 2016). No clear answer to this question has been found through empirical means, with various studies supporting either view. The issue is further aggravated by the multitude of approaches to the implementation and evaluation of CSR activities that are used and differences in the underlying frameworks with regard to what qualifies as CSR and what the goal of CSR activities should be (Garriga and Melé, 2004; Höllerer, 2012). Aggravating this issue, it is also difficult to assess the directionality of causality between CSR activities and financial performance.

Focusing on ESG performance, as previously defined, may resolve some of these problems, as data providers usually define which factors are included and how they are weighted in their ratings. Often sub-scores for the three areas, to which the acronym refers, are available, allowing separate evaluation. Generally speaking, ESG scores usually deal more with how a company operates, rather than what activities a company engages in, in addition to its primary operation. This means that a company could receive very good ESG ratings without ever engaging in any philanthropic activities.

Using ESG indicators also has the advantage that they are metrics used by investors to make investment decisions, second only to exclusion criteria in the spread of their adoption (Eccles, Kastrapeli and Potter, 2017). This means that they directly affect the market for socially responsible investment, and thus provide more practically useful insights. Results are also affected by the beliefs investors hold about using non-financial data and this may vary depending on the type of investor and their time horizon.

There is still uncertainty about what role ESG ratings play in the evaluation of a given stock. They could be indicators of quality meaning that better ESG scores would predict higher returns. At the same time this could mean that better ESG performance would be associated with higher cost of capital, especially if this relationship holds outside of equity markets.

An alternative to this would be the view that good ESG performance is associated with lower idiosyncratic risk, for example from environmental damage, legal risk and policy risk, factors unlikely to be captured by stochastic measures of volatility. This would justify investors accepting lower (risk adjusted) returns, with a reduction in non-volatility risk. It would also imply lower cost of equity for firms with better ratings, due to lower risk premiums. Unfortunately it is difficult to verify this, because measures of cost of equity are generally imprecise and difficult to substantiate (Fama and French, 1997; Gebhardt, Lee and Swaminathan, 2001), a situation that is aggravated by the increased tendency to provide shareholder return by other means than dividends. There is, however, some empirical evidence supporting the view that better ESG performance is associated firms being less capital constrained than those with worse performance (Cheng, Loannou and Serafeim, 2014).

III. FORMAL HYPOTHESIS

To resolve some of these uncertainties regarding ESG performance and returns, the following hypotheses shall be the object of this paper. As there are various and often contrary views in the literature, both the possibility of positive and negative effects are considered as separate alternative hypotheses, resulting in the following structure:

H_0: Differences in ESG performance do not predict differences in returns.

H_{A1}: Better ESG performance predicts higher returns, supporting the view that ESG measures are an indicator of quality or that highly rated stocks are preferred by investors.

H_{A2}: Better ESG performance predicts lower returns, supporting the view that they either imply reduced efficiency or reduced idiosyncratic risk.

If H_{A1} is correct, the interpretation is straightforward. If, on the other hand, H_{A2} ends up being the correct hypothesis, the mechanism of action has to be discussed. It could be the case that investors use bad ESG performance as an indicator of increased exposure to low probability, but high consequence events. These could take the form of physical damage due to environmental risk, lawsuits, or increased susceptibility negative consequences from policy change. Alternatively it could be an indicator for misapplication of funds and reduced efficiency, resulting in reduced shareholder returns.

Apart from considering just the overall ESG score, its three main areas will be evaluated seperately. They are the Environmental Pillar, the Social Pillar and the Governance Pillar. This is done, because each pillar could interact in different and possibly contrary ways with the excess returns generated by a given stock. It is quite likely that specifically the Environmental Pillar would relate to returns in a different way than the other pillars, as it captures a more technological aspect than the other components. It also covers an area that is more at risk from policy change and where changes are more costly to implement. The social and governance aspect have more to do with organizational structure and workforce management, areas where improved ratings do not necessarily require large investments and where adjustements to comply with new legal requirements are easier to implement.

IV. METHODOLOGY

4.1 Data

Two datasets have been compiled to be used in this paper. The first one combines three german stock indices. The LDAX, which includes Germanys 30 most important firms, the MDAX which comprises 60 firms ranked right below the LDAX, and the SDAX, which consists of 70 smaller firms. These were combined for a total of 180 firms and the dataset will be referred to as the DAX dataset. The second dataset, named STOXX600, comprises the members of the STOXX600 indice, which includes 600 of the most important european companies. Both datasets encompass data from the year 2000 until 2020, collected yearly at the end of the first quarter for each respective year. All low frequency variables, such as ESG scores (yearly) or enterprise reporting (quarterly) are updated within the first quarter, which means that in-period lag is minimized in comparison to high frequency variables, such as price or volatility, while still allowing time for information to diffuse and to affect market participants.

The datasets take the form of unbalanced panel data, due to companies joining the index at different times. Some datapoints, particularly ESG scores are not available for all companies due to data not having been reported at the time. Coverage for ESG data was and still is better for large companies than for smaller companies, but coverage has expanded to cover a larger number of firms over time.

To insure against the introduction of bias due to this fact, an auxilliary regression was run with the ESG score as the independent variable and market cap as the explanatory variable. While there is a clear correlation between the existence of an ESG ratings and market cap, no significant correlation between market cap and actual ESG scores has been found. Because the number of firms covered increases over time, more recent datapoints are included in the estimated models than earlier ones. This means that a certain level of recency bias is present in the estimated models, a fact that should be kept in mind. Figure 1 provides a graph indicating the ratio of firms with ESG scores present. Accordingly recent data has a somewhat higher weight in the estimated results.

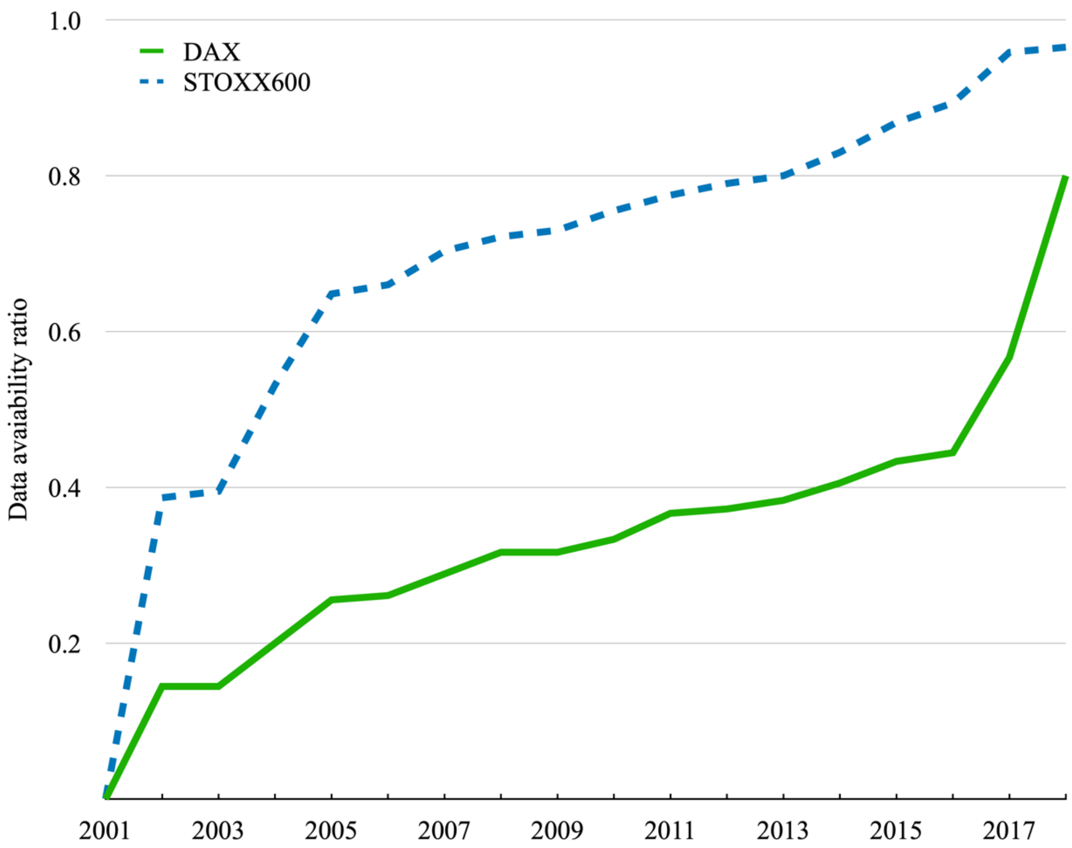

Figure 1: Data availability ratio by year (own elaboration based on Thomson Reuters Refinitiv ESG scores)

All data on individual firms used in this analysis has been collected through Thomson Reuters Datastream, which provides data sourced directly from stockmarkets, and from financial data providers such as Thomson Reuters Refinitiv and Worldscope. Data has been collected on Price, Market Value, various earnings measures, volatility, ESG scores and emissions, a variety of financial ratios, industry classification and, in the case of the STOXX600, country of headquarters.

Government bond yield were used as proxies for the risk-free interest rate. For the DAX dataset ten year german treasury bonds are used; for the STOXX600 dataset a ten year euro area benchmark rate is used. Data for both is provided by the OECD.

4.2 Selection of Variables

The following variables have been included in the models. Some have been calculated from the available data due to not being available directly.

4.2.1 Total Excess Return on Investment (RET)

Total excess return on investment over one year, calculated as seen in Equation (1). (P) is the observed price of a given stock, corrected for stock splits and similar capital measures. Dividend per share (DPS) are the expected dividend payments over the following 12 months from time t. One-off special dividend payments are not included. (RFR) is the risk free rate at time t.

$$RET_t = \frac{\Delta P_t + DPS_t}{P_{t-1}} - RFR_t \tag{1}$$

4.2.2 Total profit margin (PROF)

The profitability of a company as indicated by its total profit margin, which is calculated as indicated by Equation (2) using earnings before interest and taxes (EBIT) as well as net revenue (NREV). As Profitability (PROF) follows a power law distributions, the natural logarithm of the variable is used instead and this is denoted by l_PROF. The transformed variable is approximately normal.

$$PROF_t = \frac{EBIT_t}{NREV_t} \tag{2}$$

4.2.3 Market to Book ratio (MB)

Equation (3) indicates how the Market to book ratio was calculated using Market Value (MV) and Book Value (BV). This is a factor often used as a fundamental value indicator, and is therefore included as a controlling variable.

$$MB_t = \frac{MV_t}{BV_t} \tag{3}$$

4.2.4 β volatility indicator (HBETA)

Historical beta values for each firm as provided by Thomson Reuters Datastream. It is estimated using monthly price data for t to t-5 years for the firm and their respective index (LDAX, MDAX, SDAX or STOXX600).

4.2.5 Net Revenue Growth Rate (gNREV)

The year on year net revenue growth rate is included to account for the potential effects of changing firm size and is calculated as shown in Equation (4). Net Revenue growth rate (gNREV) was chosen over simply including Net Revenue (NREV), as an indicator of size, to avoid excessive colinearity with the unit constants, which would be present if just net revenue was used. This is the case, because firm size varies a lot between different firms and varies little over time. Changes in net revenue growth are also used by investors as an indicator of a given firms trajectory.

$$gNREV_t = \frac{\Delta NREV_t}{NREV_{t-1}} \tag{4}$$

4.2.6 ESG Scores

A number of Thomson Reuters Refinitiv Scores are used in this paper. They were selected over ESG scores by other data providers mainly due to availability and the transparency of their data generation process. While direct physical measurements for the environmental factor, such as CO_2 equivalent emissions are collected and have been used succesfully (Young In, Young Park and Monk, 2018), availability remains limited. Obvious quantifiable metrics for social and governance factors do not exist or are not sufficiently broad. Market participants also tend to use ESG scores by data providers or similar proprietary scores in their decision-making. Therefore ESG performance is approximated through ESG scores for the purposes of this paper.

The Refinitiv ESG Environmental Pillar (ESGSCE), Social Pillar (ESGSCS) and Governance Pillar (ESGSCG) each rate the performance of a company in comparison to its industry peers for each respective area. The Environmental Pillar considers factors such as resource use, emissions and technological development. The Social Pillar reflects a companies treatment of its workforce, customers and surrounding community. The Governance Pillar evaluates management compliance with corporate governance principles, equal treatment of shareholders and ESG strategy implementation. Each score is based on more than 50 metrics selected from a larger set of key performance indicators (KPI) according to industry group relevance and other factors. Scores are weighted depending on their materiality to each industry. The final scores are based on performance in comparison to industry peers for the Environmental and Social Pillars and in comparison to other firms of the same country for the Governance Pillar. The scores are normalized and distributed between 0 and 1.

Only publicly available data is used to calculate the scores. This includes financial reports, separate ESG and CSR reports, the company's website as well as information published by NGO's and news sources. Lack of published data is penalized only if it affects certain KPI's, specifically those most material to a given industry.

The Refinitiv ESG Combined Score (ESGSC) combines the three pillar scores with equal weighting and adds penalisation for ESG Controversies, such as lawsuits, imposed fines and environmental damage when reported by news media.

4.2.7 Discarded Variables

Further variables often used to assess investments in financial markets were considered, but rejected, either due to not being significant in any of the base models, presence of notable collinearity, the same information already being better captured by other included variables or lack of data. These variables were Earnings per share, Return on Assets, Return on Invested Capital, Price to Earnings ratio and Enterprise value. Dummy variables for industry and country were also discarded, as was the inclusion of a time trend.

4.3 Models

4.3.1 Transformation of the dependent variable

Total Return on Investment (RET), the dependent variable, exhibits some skew and is highly leptokurtic. This reflects a well known property of financial returns, namely fat tails (Kabašinskas *et al.*, 2009). Most often the concern is the underestimation of tail-risk in risk analysis, but it also leads to overstatement of OLS coefficients under normal (close to mean) conditions. It does not affect joint significance (Oorschot, 2017) if no skew is present, but it can lead to inefficient estimation (Stoyanov, Samorodnitsky and Rachey, 2007). A large variety of approaches exists to deal with this issue ranging from trimming the data to using robust methods such as quantile regression (Balkema and Embrechts, 2018) or working with alternative distributions for the residuals (Stoyanov, Samorodnitsky and Rachey, 2007; Kring *et al.*, 2009). For non-critical applications the issue is often addressed by winsorizing the data. In this case the divergence from normality is continouus and with some skew present, meaning that winsorizing could not resolve the issue completely. Alternative transformations such as cube-root, square-root of absolute value and logarithms were considered but did not lead to satisfactory results.

For the purposes of this paper an alternative approach has been chosen, using a two-part Lambert-W transformation which normalizes third and fourth order moments while maintaining first and second order moments, in accordance with the methodology laid out by Goerg (2015). This results in a gaussianised series named RET_G, which maintains the first- and second-order moments, a factor that aids interpretation. The method used for this involves matching two Lambert-W functions to the existing data, one for values above and another for values below the mean, using maximum likelihood estimation (Goerg, 2011, 2015). The resulting parameters (δ_l and δ_r) can then be used with the inverse of the Lambert-W function to remove skew and fat tails from the data.

Applying the non-inverse function to results of models using this transformation allows producing point estimates. Generating a partial derivative of the function provides scaling factors that can be applied to predictions made by the model directly.

$$Y = (X - \mu_x) * e^{\frac{\delta}{2} * \left(\frac{X-\mu_x}{\sigma_x}\right)^2} + \mu_x \tag{5}$$

$$\frac{\partial Y}{\partial X} = e^{\frac{\delta}{2} * \left(\frac{X-\mu_x}{\sigma_x}\right)^2} + (X - \mu_x) * e^{\frac{\delta}{2} * \left(\frac{X-\mu_x}{\sigma_x}\right)^2} * \frac{\delta}{2} * 2 * \frac{1}{\sigma_x} * \left(\frac{X - \mu_x}{\sigma_x}\right) \tag{6}$$

Equation (5) represents the underlying Lambert-W function with with X being the gaussianised data and Y being the original data, both taking the form of a N*T matrix. δ would be replaced by δ_l and δ_r depending on which side of μ_x the values are. Equation (6) is the derivative of (5) with respect to X and can be used to calculate scaling factors.

4.3.2 Models

A variety of estimators were considered for this analysis. The primary candidates were simple OLS estimation, both random-effect and fixed-effect panel estimators as well as dynamic panel models. Dynamic models were rejected as no significant amount of autocorrelation could be detected.

In the end fixed-effect OLS estimation was chosen, as this is indicated by the Joint significance test for differing group means when compared to pooled models, even when including dummies for Industry classification and country of origin, with p-values below 0.01 for both the DAX and STOXX600 datasets. The Breusch-Pagan test indicates that a pooled model is more appropriate than a random effects model, failing to reject H_0. The Hausman test indicates that a fixed-effects model is a better fit than the random effects alternative, again with p-values below 0.01.

Two main models were created and estimated on a variety of samples. Model 1 takes the form described in Equation (7) and was intended to measure the overall effect of ESG-Performance on stock returns while controlling for a number of other variables.

$$RET\_G_{i,t} = \beta_0 c_i + \beta_1 ESGSC_{i,t} + \beta_2 HBETA_{i,t} + \beta_3 l\_PROF_{i,t} + \beta_4 l\_MB_{i,t} + \beta_5 d\_NREV_{i,t} \tag{7}$$

Model 2 replaces the ESG Combined score with a separate score for each pillar, to assess if there are differences in how each area relates to excess returns. The same controling variables as in Model 1 are included. Model 2 is defined by Equation (8).

$$RET\_G_{i,t} = \beta_0 c_i + \beta_1 ESGSCE_{i,t} + \beta_2 ESGSCE_{i,t} + \beta_3 ESGSCE_{i,t} + \beta_4 HBETA_{i,t} + \beta_5 l\_PROF_{i,t} + \beta_6 l\_MB_{i,t} + \beta_7 d\_NREV_{i,t} \tag{8}$$

Apart from models using gaussianised returns (RET_G), both non-modified (RET) and winsorized returns (RET_W) were evaluated to insure the integrity of the results. As expected coefficients were smaller for winsorized and gaussianised data than for unmodified data and

R^2 better for gaussianised returns than for the winsorised data. Accordingly gaussianised returns were used for all subsequent evaluations.

The same models were run on restricted datasets to evaluate the presence of structural breaks and to see if the relationships between variables have changed over time. The period during and just after the great recession was also evaluated separately, to test if good ESG performance predicts higher robustness in a downturn.

The following models were run on restricted datasets to see if the predicted variable relationships differ for specific time periods:

1. Model E1 and E2 consider the period from 2001 to 2008, the period before the financial crisis.
2. Model R1 and R2 consider the period from 2012 up till 2020, to test if the notable increase in Institutional investors taking into account ESG criteria as well as the emergence of specific ESG funds has lead to significantly different coefficients.
3. Model CR1 and CR2 only take into account the period from 2008 to 2009 to see how the variables would interact during times of economic crisis.

Additionaly, differences between firms based on their size were to be evaluated. While size differences were already corrected for by the base variables, additional models with longitudinal restrictions were run to test if coefficients would vary between groups. In this case, size would act as a multiplicative factor for ESG scores or other variables. In the case of DAX dataset the following subsamples were evaluated:

1. Model L1 and L2 only take into account LDAX constituents, the 30 largest public companies in Germany.
2. Model M1 and M2 are limited to MDAX constituents – a total of 80 firms right below the LDAX in term of size.
3. Model S1 and S2 are limited to members of the SDAX – a total of 70 firms, ranking right below the MDAX.

For the STOXX600 a different approach was chosen to classify firms as large-, mid- and small-caps, which were then used to run restricted models:

1. Model L1 and L2 are limited to companies with market caps above the 50th percentile of the STOXX600 indice. This is equivalent to a market value of about €10 billion in 2019, which is a commonly chosen cutoff point between large- and mid-caps. It is also the cutoff used by various EUROSTOXX mid-cap indices.
2. Model MS1 and MS2 take into account all companies below the 50th percentile in market-cap. Mid- and small-caps were not seperated, because the number of small-caps in the STOXX600 is very limited. When using the commonly used cutoff of about €2 billion in 2019, which is also used by EUROSTOXX mid-cap indices, the cutoff for small-caps would be below the 5th percentile.

V. RESULTS

If not denoted otherwise, α = 0.05 is used as the relevant level of significance. Within the tables that summarize results, stars are used to indicate significance. One star refers to $\alpha < 0.1$, two stars refer to $\alpha < 0.05$ and three stars refer to $\alpha < 0.01$, in accordance with the convention of many statistical software packages.

5.1 Results from Gaussianization

Table 1: DAX Lambert-W coefficient estimates

μ_x	0.1158
σ_x	0.3472
δ_l	0.0150
δ_r	0.1617

Table 1 indicates the results of the auxilliary regression used to gaussianise the returns and therefore also the coefficients to be used for back-transformation. Figure 2 provides a graphical representation of the adjustement function and it's first partial derivative, in correspondence with Equations (5) and (6) respectively.

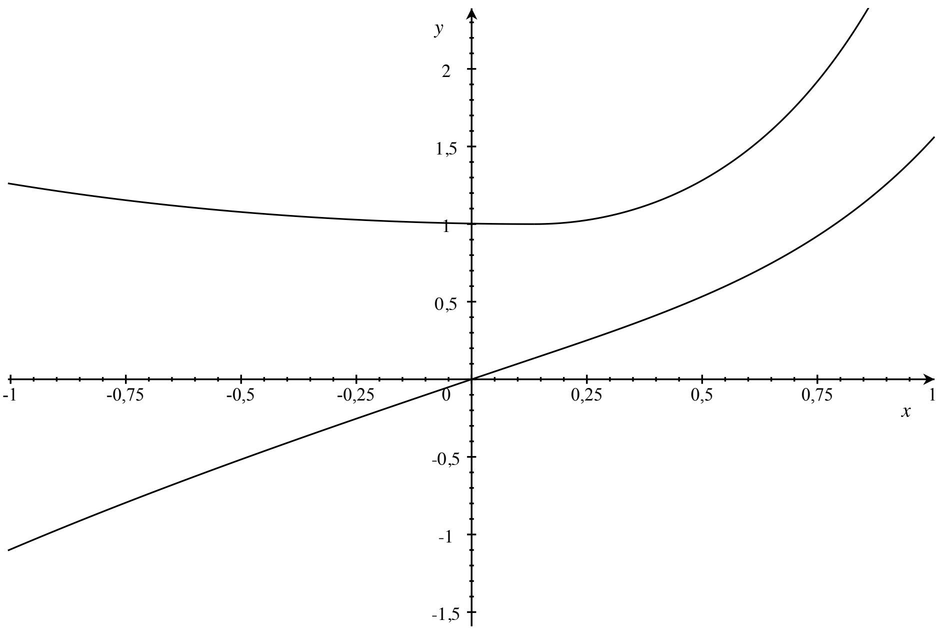

Figure 2: Lambert-W adjustment function (bottom) and partial derivative (above)

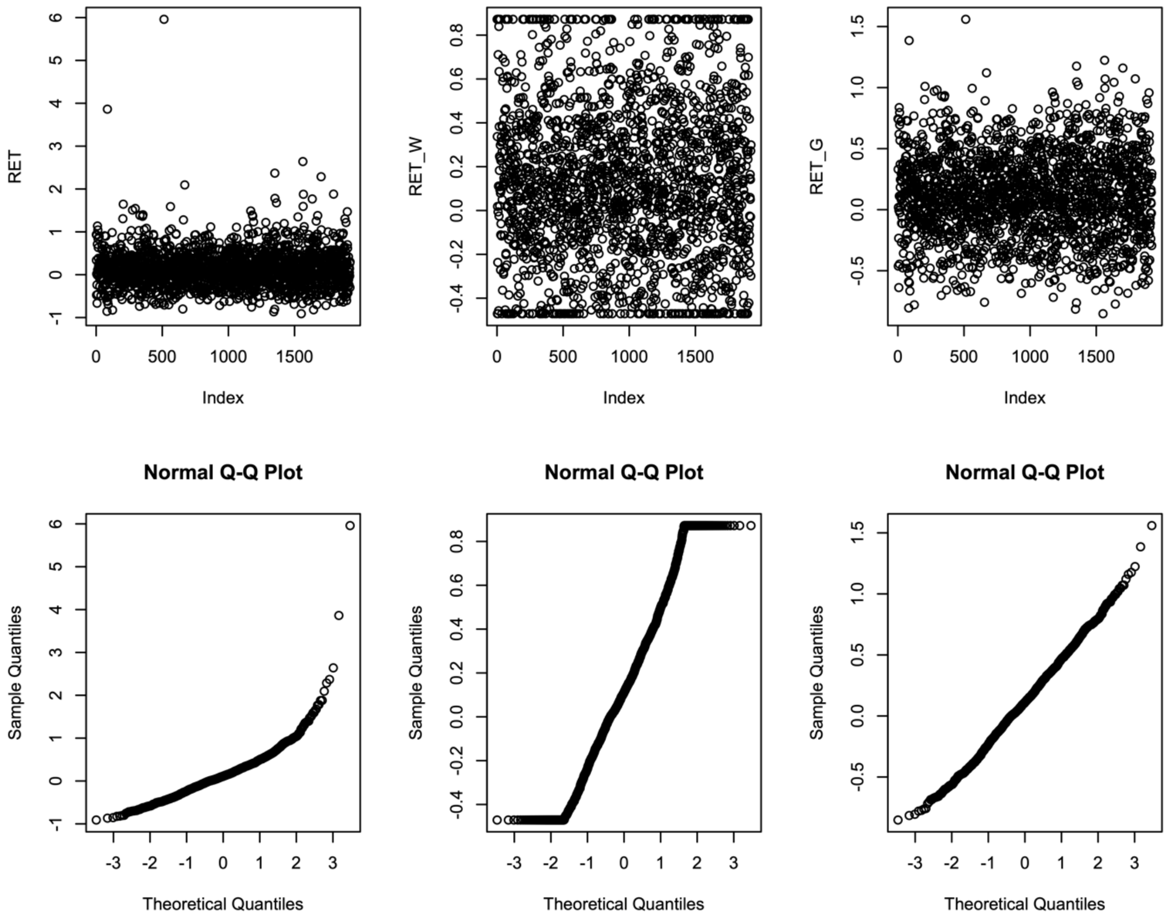

Figure 3: Sample and Q-Q plots for RET, RET_W and RET_G

Figure 3 provides sample plots and Q-Q plots for the original series (RET), a winsorized series with two-sided 5% cutoffs (RET_W) and for the gaussianised series (RET_G), for comparison purposes. Table 2 provides an overview for the summary statistics of the same series. Under normal circumstances (close to mean), predictions based on RET_G do not have to be adjusted, For predictions further away from the mean, adjustements would be recommended.

Table 2: DAX Summary statistics for the dependent variable

	RET	**RET_W**	**RET_G**
Mean	0.1464	0.1335	0.1187
Median	0.1114	0.1114	0.1114
Std. Dev.	0.4333	0.351	0.3456
Skewness	2.2479	0.2561	0.1016
Ex. kurtosis	20.925	-0.4886	-0.0001

5.2 Base Models

Table 3: DAX Base Model 1 Results

```
Model 1: Fixed-effects, using 881 observations
Included 124 cross-sectional units
Time-series length: minimum 1, maximum 18
Dependent variable: RET_G

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       0.545763     0.0762835     7.154    2.00e-12 ***
  ESGSC      -0.151178     0.0836980    -1.806    0.0713   *
  HBETA      -0.0750875    0.0374106    -2.007    0.0451   **
  l_PROF      0.148303     0.0213034     6.961    7.34e-12 ***
  MB          0.0814808    0.0190702     4.273    2.18e-05 ***
  gNREV       0.367805     0.0715683     5.139    3.52e-07 ***

Mean dependent var   0.137151   S.D. dependent var   0.318288
Sum squared resid    65.55982   S.E. of regression   0.295264
LSDV R-squared       0.264616   Within R-squared     0.150607
LSDV F(128, 752)     2.114023   P-value(F)           7.03e-10
Log-likelihood      -105.6242   Akaike criterion     469.2483
Schwarz criterion    1086.005   Hannan-Quinn         705.0947
rho                 -0.182900   Durbin-Watson        1.946220
```

Table 3 provides the results of the first base model, taking into account the overall ESG score and a number of other independent variables. The model is jointly significant and results in a LSDV goodness of fit of 0.26. All coefficients are significant at the selected level of significance, apart from the ESG score with a p-value of 0.07. The ESG score is negatively correlated with the dependent variable, supporting the second alternate hypothesis H_{A2}.

In this case a one percentage point increase in the ESG score is correlated with a 0.15 p.p. decrease in returns. The coefficients of all other variables are in line with intuition and previous results. A one percent increase in profitability results in a 0.15 p.p. increase in returns. A one p.p. increase in the market to book ratio is correlated with a 0.08 p.p. increase in returns. Similarly a one p.p. increase in the net revenue growth rate is associated with a 0.37 p.p. increase in returns. Higher volatility as measured by historical β values are correlated with lower returns, though this might be the result of including the financial crisis in the dataset, which had an outsize impact on both volatility and returns.

Table 4: DAX Base Model 2 Results

Model 2: Fixed-effects, using 881 observations
Included 124 cross-sectional units
Time-series length: minimum 1, maximum 18
Dependent variable: RET_G

	coefficient	std. error	t-ratio	p-value	
const	0.372618	0.0881019	4.229	2.63e-05	***
ESGSCE	−0.286081	0.101171	−2.828	0.0048	***
ESGSCS	0.273632	0.101413	2.698	0.0071	***
ESGSCG	0.178029	0.0829463	2.146	0.0322	**
HBETA	−0.0737674	0.0374909	−1.968	0.0495	**
l_PROF	0.144901	0.0212190	6.829	1.77e-11	***
MB	0.0823914	0.0189605	4.345	1.58e-05	***
gNREV	0.356728	0.0711491	5.014	6.66e-07	***

Mean dependent var	0.137151	S.D. dependent var	0.318288
Sum squared resid	64.62551	S.E. of regression	0.293543
LSDV R-squared	0.275096	Within R-squared	0.162712
LSDV F(130, 750)	2.189384	P-value(F)	7.98e-11
Log-likelihood	−99.30133	Akaike criterion	460.6027
Schwarz criterion	1086.921	Hannan-Quinn	700.1055
rho	−0.201689	Durbin-Watson	1.975412

Table 4 provides the result of the second base model, in which the three ESG pillars are included separately. The model shows a LSDV R^2 of 0.28, with all explanatory variables being significant. The Environmental Pillar is negatively correlated with returns with a one p.p. increase in the score correlated with a 0.32 p.p. reduction in returns. The opposite is true for the Social Pillar and Governance Pillar where a one pp increase is correlated with a 0.23 p.p. and 0.17 p.p. increase, respectively. All additional variables are in line with the results observed in Model 1. Collinearity between the different ESG scores is found to be limited, though there is some collinearity with the unit dummys present according to the Belsley-Kuh-Welsch collinearity diagnostics. To insure against erroneous induction resulting from interaction between the three ESG pillar scores, three additional models were erstimated, each omitting two of the ESG pillar scores. The resulting coefficients were all close to those estimated in Model 2, with no coefficients diverging to a significant extend.

The results of Model 2 indicate that for the Environmental Pillar H_{A2} is valid and that the variable is, therefore, associated with lower shareholder return. The social and Governance Pillar scores both have significant and positive coefficients, implying that H_{A1} is valid and that they are indicators of quality.

Both models do not show significant first-order autocorrelation according to the Durbin-Watson and Wooldridge tests. Seasonal effects can be discounted as yearly data is used. The residuals are approximately normal, though they fail the usual normality tests. Both skew and kurtosis of the residuals are very limited, with a mean very close to zero. Serial correlation and heteroskedacity could also not be detected, resulting in spherical errors and, therefore, unbiased, consistent and efficient estimations. Both models were also run using the non-gaussianised returns. Directionality of results is the same for all variables though coefficients are slightly larger, with notably increased standard errors in accordance with the expected results of using non-gaussianised data. Residuals are also notably less normally distributed and much more leptokurtic.

5.3 Evaluation of structural change

Table 5: DAX results for different time periods

	Model E1		**Model R1**		**Model CR1**		**Model E2**		**Model R2**		**Model CR2**	
coefficients	*2001-2008*		*2012-2020*		*2008-2009*		*2001-2008*		*2012-2020*		*2008-2009*	
const	0.6032	***	0.3565	***	0.4292		0.3792	**	0.4071	**	2.6760	*
ESGSC	-0.0446		-0.2408	**	-0.6108							
ESGSCE							-0.0221		-0.4049	**	-1.4900	
ESGSCS							0.2637		0.1403		-2.2710	
ESGSCG							0.1451		0.0368		-0.4232	
HBETA	0.0873		-0.0914		-0.3896	*	0.0593		-0.1073	*	-0.2650	
l_PROF	0.2467	***	0.0768	***	-0.0231		0.2440	***	0.0834	***	-0.0426	
MB	0.1048	**	0.2191	***	0.0988	*	0.1142	***	0.2268	***	0.1083	**
gNREV	0.4357	**	-0.0393		0.9144	*	0.3774	**	-0.0451		1.2490	**
p-values												
Joint	0.0895	*	0.0000	***	0.0001	***	0.0669	*	0.0000	***	0.0002	***
const	0.7482		0.0728	*	0.8027		0.9725		0.8393		0.1419	
ESGSC	0.5173		0.3887		0.3696							
ESGSCE							0.1372		0.4679		0.4954	
ESGSCS							0.9522		0.5035		0.0674	*
ESGSCG							0.8178		0.3120		0.2858	
HBETA	0.1285		0.7735		0.1252		0.2117		0.5701		0.3812	
l_PROF	0.0470	**	0.0141	**	0.1285		0.0436	**	0.0362	**	0.0980	*
MB	0.5911		0.0016	***	0.7459		0.4633		0.0011	***	0.6270	
gNREV	0.6858		0.0000	***	0.2448		0.9023		0.0000	***	0.0724	*
n	235		499		90		235		499		90	
LSDV R^2	0.41		0.35		0.66		0.41		0.35		0.69	

Table 5 summarizes the results of Models 1 and 2 being run on restricted time periods to evaluate if the underlying behaviour has changed over time. The upper half of the table denotes the estimated coefficients with their respective level of significance. All models were tested by using the base model coefficients as linear restrictions for the restricted models and testing for both joint and individual significance. The resulting p-values are shown in the lower half of the table. The linear restriction tests for each variable are equivalent to the usual t-test, but using the previously estimated base model coefficients in place of zero.

Overall, the relationships between the variables have remained stable over time in most cases, with a few noteworthy exceptions. Jointly the models restricted to recent data and those restricted to the financial crisis period are significantly different from the base models. None of the ESG scores differ to a significant extend from their base model values. Despite that some observations can be made: In the case of Model 1, the effect of ESG scores during the pre-crisis period is negative, but much smaller and not significantly different from zero. For recent data the effect is slightly larger than in the base model, but not significantly different from it. It is however different from the early period (p-value = 0.059), indicating that a change in this relationship has occurred.

Regarding the control variables, it can be observed that historical β values have very limited effect when ignoring the financial crisis, neither one being significantly different from zero. The positive correlation between profitability and returns is notably smaller in the recent data than it was the case in the period before the financial crisis. The same is true for net revenue growth, which recently are not significantly different from zero, while being significantly different from the base models. At the same time the positive coefficient for a given stocks market to book ratio has increased notably and also differs significantly from the base models for the more recent subsample.

For Model 2 similar observations can be made. Behaviour of the control variables is similar to that observed for Model 1. The effect of the Environmental Pillar score is smaller and not significant for the early period, while being much stronger and significant for the more recent period. At the same time the, positive effects for the social and Governance Pillar are larger in the earlier period and notably smaller in the later period, though none of these coefficients are significant in their own right. This tendency could be explained by more widespread consideration of these factors as indicators of reduced risk and, therefore, requiring lower returns. The lower levels of significance observed for the early period likely is the result of the diminished number of valid observations, with only a limited number of firms with ESG scores.

When observing the period during the financial crisis, both models show strong negative correlations for both the overall ESG score as well as all ESG sub-scores, though none of these results are significant. When taken at face value, the behaviour would be explained if we assume that better ESG performance usually requires large fixed-cost expenses and less flexible employment models. It has to be kept in mind, however, that none of the values are significantly different from the base models or from zero.

5.4 Differences between firms of different size

Table 6: DAX results by index membership

	Model L1		Model M1		Model S1		Model L2		Model M2		Model S2	
coefficients	*LDAX*		*MDAX*		*SDAX*		*LDAX*		*MDAX*		*SDAX*	
const	0.4619	***	0.5242	***	0.9037	**	0.2449	**	0.4699	**	0.5140	
ESGSC	-0.1015		-0.3696	**	0.1919							
ESGSCE							-0.4017	***	-0.1989		-0.1298	
ESGSCS							0.4857	***	0.0153		0.2161	
ESGSCG							0.1441		-0.0873		0.7132	***
HBETA	-0.0625		-0.0693		-0.1366		-0.0915	*	-0.0712		-0.0764	
l_PROF	0.1447	***	0.1254	***	0.2040	***	0.1316	***	0.1213	***	0.1755	***
MB	0.1045	***	0.1749	***	0.0350		0.1118	***	0.1813	***	0.0453	
gNREV	0.4744	***	0.6329	***	0.1401		0.4558	***	0.6357	***	0.1169	
p-values												
Joint	0.0669	*	0.0000	***	0.0002	***	0.1327		0.2619		0.1346	
const	0.9725		0.8393		0.1419		0.2698		0.6150		0.7121	
ESGSC	0.1372		0.4679		0.4954							
ESGSCE							0.4140		0.6357		0.5334	
ESGSCS							0.1103		0.1658		0.8364	
ESGSCG							0.7502		0.1090		0.0191	**
HBETA	0.9522		0.5035		0.0674	*	0.7288		0.9707		0.9792	
l_PROF	0.8178		0.3120		0.2858		0.6454		0.5909		0.5544	
MB	0.2117		0.5701		0.3812		0.3673		0.1066		0.2250	
gNREV	0.0436	**	0.0362	**	0.0980	*	0.4283		0.0645	*	0.0285	**
n	392		308		181		392		308		181	
LSDV R^2	0.21		0.24		0.43		0.24		0.23		0.48	

Table 6 shows the results of Model 1 and 2 with the dataset restricted by index membership. Only Model M1 and S1 are significantly different from the base model. It also is necessary to take note of the reduced number of observation included in the models for the MDAX and SDAX, despite those indices including a much larger number of companies than the LDAX. This is caused by the significantly worse coverage for constituents of those indices, particularly with regard to ESG scores, which leads to fewer valid obervations.

For Model 1 the negative effect of the overall ESG score is larger and more significant for MDAX constituents than for either the LDAX or SDAX, but none of the results differ to a significant degree from the base model. Most of the controlling variables are significant and in in line with the base model results. The only exception to this in the net revenue growth rate which is significantly larger for LDAX and MDAX members than for smaller firms, with those two coefficients being significantly different from the base models.

For Model 2 the estimated coefficients for the Environmental Pillar and Social Pillar are notably larger and more significant for large firms than for MDAX and SDAX constituents. None of the scores apart from the governance score in the case of SDAX firms are significantly different to the base model. The SDAX Governance Pillar score is notably larger than for other firms, the estimated coefficient being significantly different from the base model and from zero.

Taking a more general prespective it seems that the estimations of the base model apply to firms largely independent of their size. While there are significant differences in some of the variables, these differences are limited to magnitute. The directionality of the results is maintained in all cases.

5.5 Comparison with the STOXX600 Dataset

The same estimations that were performed on the DAX dataset were repeated on the STOXX600 dataset to corroborate the results of the former with a set of firms working under more heterogeneous circumstances and to see if the results could be generalized. When comparing base model results, it can be noted, that directionality has been maintained in all cases and the resulting coefficients follow a similar pattern as the DAX models. There are however some differences with regard to magnitude and significance of some of the variables. In general, the STOXX600 models resulted in lower R^2 values, a result likely explained by the fact that companies are operating under much more varied circumstances. Members of the

STOXX600, smaller ones in particular, not only serve different primary markets, they also operate under differing regulatory and tax regimes. While fixed-effect estimators are able to account for country and industry fixed effects, they are not able to account for differing development between countries over the period of time that has been considered.

Table 7: STOXX600 Lambert-W coefficient estimates

μ_x	0.1008
σ_x	0.2837
δ_l	0.0483
δ_r	0.1543

The estimated factors used to gaussianise the returns (Table 7) are very similar to those used for the DAX returns, indicating that returns for the DAX and STOXX600 datasets follow a similar distribution. Table 8 and Table 9 show the base model results for the STOXX600 dataset.

Table 8: STOXX600 Base Model 1 results

```
Model 1: Fixed-effects, using 6160 observations
Included 550 cross-sectional units
Time-series length: minimum 1, maximum 18
Dependent variable: RET_G

             coefficient    std. error    t-ratio   p-value
  ---------------------------------------------------------
  const       0.369852      0.0235587     15.70     2.18e-54 ***
  ESGSC      -0.0127362     0.0274149     -0.4646   0.6423
  HBETA      -0.0374136     0.0115272     -3.246    0.0012   ***
  l_PROF      0.110202      0.00699954    15.74     1.10e-54 ***
  MB          0.00537957    0.00165857     3.243    0.0012   ***
  gNREV       3.75829e-05   4.95863e-05    0.7579   0.4485

Mean dependent var   0.114725   S.D. dependent var   0.271660
Sum squared resid    396.3296   S.E. of regression   0.265914
LSDV R-squared       0.128046   Within R-squared     0.047434
LSDV F(554, 5605)    1.485719   P-value(F)           1.44e-11
Log-likelihood      -290.4171   Akaike criterion     1690.834
Schwarz criterion    5423.671   Hannan-Quinn         2985.414
rho                 -0.108121   Durbin-Watson        1.962985
```

Table 9: STOXX600 Base Model 2 results

```
Model 2: Fixed-effects, using 6160 observations
Included 550 cross-sectional units
Time-series length: minimum 1, maximum 18
Dependent variable: RET_G

             coefficient    std. error    t-ratio   p-value
  ---------------------------------------------------------
  const       0.349005      0.0297836     11.72     2.39e-31 ***
  ESGSCE     -0.0993420     0.0332182     -2.991    0.0028   ***
  ESGSCS      0.110891      0.0344201      3.222    0.0013   ***
  ESGSCG      0.00936190    0.0260223      0.3598   0.7190
  HBETA      -0.0358059     0.0115324     -3.105    0.0019   ***
  l_PROF      0.109224      0.00699676    15.61     8.27e-54 ***
  MB          0.00536906    0.00165935     3.236    0.0012   ***
  gNREV       3.14692e-05   4.95720e-05    0.6348   0.5256

Mean dependent var   0.114725   S.D. dependent var   0.271660
Sum squared resid    395.4026   S.E. of regression   0.265650
LSDV R-squared       0.130085   Within R-squared     0.049662
LSDV F(556, 5603)    1.506942   P-value(F)           2.34e-12
Log-likelihood      -283.2045   Akaike criterion     1680.409
Schwarz criterion    5426.697   Hannan-Quinn         2979.654
rho                 -0.109903   Durbin-Watson        1.966662
```

In the case of Base Model 1, the ESG score coefficient is only slightly negative, and not significant. The same is true for the net revenue growth rate. The other variables vary in magnitude but are significant and in line with the DAX results. Base Model 2 results are more closely related to the DAX results. The Environmental Pillar score is significant and negatively correlated with returns, though to a smaller extend than in the DAX model. The positive effect of the Social Pillar is also significant and smaller than in the DAX dataset. The Governance Pillar score, while positive, is close to zero and not significant. The other variables are all smaller but similar to the DAX model, with only net revenue growth being not significant.

Table 10 shows the results of estimating different subsamples to check for structural change. Overall the results are again similar to the DAX dataset, though smaller in magnitude. Differences from the base model are limited, particularly for the ESG scores. In the case of Model 1 the overall ESG score is close to zero and not significant for both the early and the more recent subsamples. Profitability is in line with the DAX model with a reduced effect in the recent subsample. Interestingly higher volatility is significant and positive for the early period. It is worth noting that the net revenue growth rate is not significantly correlated with returns, apart from during the financial crisis.

Table 10: STOXX600 results for different time periods

coefficients	Model E1 2001-2008		Model R1 2012-2020		Model CR1 2008-2009		Model E2 2001-2008		Model R2 2012-2020		Model CR2 2008-2009	
const	0.3123	***	0.2287	***	0.4305	***	0.2383	***	0.2514	***	0.8985	***
ESGSC	0.0091		-0.0248		-0.3269	**						
ESGSCE							-0.0418		-0.1038		-0.4955	**
ESGSCS							0.1058	**	0.0928		-0.3471	
ESGSCG							0.0708		-0.0422		-0.1498	
HBETA	0.0822	***	0.0164		-0.3306	***	0.0789	***	0.0141		-0.3205	***
l_PROF	0.1569	***	0.0534	***	0.1228	***	0.1547	***	0.0528	***	0.1212	***
MB	0.0549	***	0.0022		0.2256	***	0.0552	***	0.0023		0.2139	***
gNREV	0.0001		0.0027		0.2760	***	0.0001		0.0021		0.2759	***
p-values												
Joint	0.0000	***	0.0000	***	0.0000	***	0.0000	***	0.0000	***	0.0000	***
const	0.2421		0.0000	***	0.6065		0.0546	*	0.0902	*	0.0226	**
ESGSC	0.6651		0.7437		0.0336	**						
ESGSCE							0.2615		0.9475		0.1095	
ESGSCS							0.9228		0.7641		0.0800	*
ESGSCG							0.1643		0.2259		0.3168	
HBETA	0.0001	***	0.0027	***	0.0000	***	0.0001	***	0.0058	***	0.0000	***
l_PROF	0.0005	***	0.0000	***	0.7070		0.0008	***	0.0000	***	0.7214	
MB	0.0000	***	0.0705	*	0.0000	***	0.0000	***	0.0807	*	0.0000	***
gNREV	0.5468		0.6779		0.0000	***	0.5311		0.7457		0.0000	***
n	1928		3170		692		1928		3170		692	
LSDV R^2	0.33		0.18		0.59		0.33		0.18		0.59	

For Model 2 none of the ESG sub-scores are significantly different from the base models. The coefficients for the Environmental Pillar are notably more negative recently than they were before the great recession, but they are not significantly different from zero or the base model. The Social Pillar score is mostly in line with the base model. The Governance Pillar is not significant for either period, as it was the case for the base model. The controling variables are in line with the results observed for Model 1.

During the financial crisis a similar pattern to the DAX dataset can be observed. The overall ESG score has a notably more negative score than under normal circumstances and is significantly different from the base model. With regard to Model 2, all ESG sub-scores are negatively correlated with total returns, as was the case in the DAX dataset. The Environmental Pillar score is significant, while the other pillar scores are negative, but not significant. None of the values are significantly different from the base models, however.

Table 11: STOXX600 results for large-caps and smaller firms

	Model L1		Model MS1		Model L2		Model MS2	
coefficients	*Large-Caps*		*Small- & Mid-Caps*		*Large-Caps*		*Small- & Mid-Caps*	
const	0.1331		0.2838	***	0.1119		0.3358	***
ESGSC	0.0200		-0.0712					
ESGSCE					-0.0368		-0.1855	***
ESGSCS					0.1413	***	0.0816	
ESGSCG					0.0379		-0.0317	
HBETA	-0.0238		-0.0378	**	-0.0273	*	-0.0360	*
l_PROF	0.1091	***	0.1090	***	0.1080	***	0.1067	***
MB	0.0075	**	0.0043	**	0.0073	***	0.0047	**
gNREV	0.0000		0.0068		0.0000		0.0060	
p-values								
Joint	0.3314		0.6934		0.2231		0.3553	
const	0.0241	**	0.4328		0.0248	**	0.9053	
ESGSC	0.3304		0.2231					
ESGSCE					0.1541		0.1054	
ESGSCS					0.5092		0.5887	
ESGSCG					0.4047		0.3143	
HBETA	0.3791		0.9853		0.5838		0.9938	
l_PROF	0.9066		0.9141		0.8924		0.8185	
MB	0.4415		0.6207		0.4818		0.7435	
gNREV	0.8413		0.3200		0.8477		0.3782	
n	3457		2703		3447		2703	
LSDV R^2	0.15		0.17		0.15		0.18	

The results of estimating coefficients for different sizes of firms are mostly similar to those resulting from the DAX dataset. None of the subsample models are jointly different to their respective base models. The same is true for all individual variables when ignoring the constant terms.

For Model 1 the negative correlation between ESG score and returns is not found for large firms. For mid- and small-caps a negative relationship is found, but it is not sufficiently significant. The coefficients estimated for the controlling variables are in line with the base Model, significant and stable. A similar pattern can be observed for Model 2. Particularly the Environmental Pillar score is more negative for smaller firms than for larger firms, and also more significant. It is also notably different from the base model, though only with a p-value of 0.099. The Social Pillar score is larger and more significant for large firms than for smaller firms, while the Governance Pillar is small and not significant in any case.

These results are not too surprising when comparing the structure of the DAX and STOXX600 datasets. In the DAX dataset only 25 firms had market-caps above € 10bn in 2019. In the STOXX600 about half of the firms are above that cutoff point. In that sense the DAX dataset more closely resembles the mid- and small-cap dataset of the STOXX600, than the set for large-caps. The majority of the LDAX would rank above, but close, to the 50th percentile of the STOXX600.

VI. DISCUSSION

6.1 Discussion of results

Overall the analysis described in this paper confirms that ESG scores are significantly related to the returns a given stock achieves. Over the last two decades a higher combined ESG score is associated with slightly lower returns for DAX firms. While the same relationship can be observed for the STOXX600, results were smaller and not significant enough to be considered confirmatory.

More interesting are the results of evaluating the relationship between the three ESG pillar scores and returns. Notably the relationship is different for the different components of the ESG score. This is a result that warrants some discussion.

The Environmental Pillar score is negatively correlated with returns. This result was found to be valid in both datasets, even though the effect size was smaller in the case of the STOXX600, meaning that hypothesis H_{A2} can be assumed to be generally valid. The existing literature may provide some insight into which of the potential mechanisms — reduced efficiency or reduced risk — is the actual mechanism causing this result. Assuming that lower risk is the correct mechanism, capital should be available relatively easily to firms with good environmental records. This is a result confirmed in a recent study by Cheng, Loannou, and Serafeim (2014), which confirms that firms with better ESG performance, particularly with regard to the Environmental Pillar, are less capital constrained. A different study employing a Fama-French model approach (Young In, Young Park and Monk, 2018), finds that firms with lower carbon emissions (one of the most important factors in the Environmental Pillar score) tend to have higher risk adjusted returns in models that include additional risk indicators apart from volatility. Another study that analysed socially responsible mutual funds (Chatterjee *et al.*, 2018) also confirms this hypothesis. Additionaly survey results indicate that asset managers use ESG ratings for the purposes of risk management (van Duuren, Plantinga and Scholtens, 2016).

Another potential indicator to assess the mechanism of this result is by considering how the environmental scores effect has changed over time. If the observed relationship is the result of less efficiency and wasted funds, one would assume that it would lead to reduced shareholder returns consistently. If it was however the result of being seen as an indicator of risk, a change in behaviour after the financial crisis would be expected, as low returns in the bond markets have made stocks more attractive to traditionally low risk investors. Increased interest in environmental concerns and sustainable investment would also lead to increased effect size if investors take these factors into account. This is a result clearly observed in the data, where the estimated coefficient is much larger for recent data than for data from before the financial crisis. While this should not be seen as proof, it can be considered as convincing evidence, particularly in consideration of the additional literature mentioned.

The results for the Social- and Governance Pillar, which indicate a positive relationship with returns allow a somewhat more straightforward explanation as being indicators of good workforce and supply chain relations and good governance. Particularly in the case of the governance score, the relative importance of shareholder rights and management quality indicators means that a positive effect matches intuition. These results are also in line with the literature, where studies that are mostly concerned with the social and governance aspects of CSR find that better scores are associated with less agency concerns and better governance (Ferrell, Liang and Renneboog, 2016) and lower information asymetry (Cui, Jo and Na, 2018).

While the positive effect for the Governance Pillar could not be confirmed with the STOXX600 dataset, this could be the result of STOXX600 members being mostly homogenous due to their size, but the score using country specifc peer groups. As large international companies tend to suffer less from governance concerns than smaller firms, STOXX600 members may all rate highly compared to their national peer group, leading to similar scores for most companies and obfuscating differences between STOXX600 constituents. For the DAX dataset, where all firms are part of the same peer group, this issue is not present. It can, therefore, be assumed valid that H_{A1} is the correct hypothesis for the governance and social components of ESG performance, meaning that these metrics are indicators of quality.

When considering how estimated coefficients change for different time periods, it is quite obvious that particularly the Environmental Pillar is increasingly an important measure to be considered. For firms, it means that bettering environmental metrics may be a net benefit due to reducing cost of equity. To some extend these improvements could also carry over to reduce cost of debt as well (*The Economist*, 2020), thereby reducing cost of capital in general. For investors it means that, increasingly, concern about policy risk and changing capital allocation have to be taken into account when determining the risks and potential rewards of an investment.

When looking at different sizes of firms, the results seem to apply more strongly to small large-caps and large mid-caps. This is likely the case, because firms in this group vary the most with regard to how they approach ESG issues. Most truly large firms already have advanced strategies in place, while too little data is available for small-caps. These results can be expected to change however, as coverage by third-party data providers expands to an increasing number of firms. For truly small public firms that are not traded regularly in liquid markets, the benefits of adopting ESG strategies are likely limited and might not cover the incurred costs. The specific industry a firm operates in, can also affect how costly collecting data on ESG performance is and if improving those metrics is a worthwhile endeavour. Trying to enforce CSR activities from a policy side has been shown to negatively affect shareholder value and waste funds, especially if crudely implemented, while the same is not true for voluntary activities (Manchiraju and Rajgopal, 2017).

6.2 Potential for further inquiry

There are a number of further avenues of research that could expand on the results presented in this paper. The most direct ways of doing so would be to confirm if these effects are valid when using ESG scores by other providers and how the results vary depending on the methodolody employed by the data provider. Another area of interest would be to test if the results of this paper are valid for other markets, particularly the US market and Asian markets, as both regulation and the attitudes of market participants may differ considerably in comparison to Europe. In addition to that, a similar model to the one used here could be applied to bond markets as cost of debt is much easier to measure than cost of equity. This would likely allow better discernment of risk and return. Apart from that, it is quite likely that future EU legislation, especially the EU Green Finance Initiative and the European Green Deal could provide natural experiments that could evaluated to test if better environmental scores did indeed protect against policy risk.

VII. CONCLUSION

This paper provides some clear evidence that better environmental performance is associated with lower returns, but likely lower idiosyncratic risk as well, while both better social and better governance performance can be seen as indicators of quality that are associated with higher returns. This divergence in the effect of the disaggregated ESG measures indicates that they describe factors that, altough often used together, interact with the potential risks and returns of a given investment in distinct ways. Overall the findings indicate that ESG ratings can be employed productively by investors. For firms they mean that improving certain areas of ESG performance can be used to reduce cost of capital. Extrapolating current trends would mean that these issues will only become more important over time.

VIII. BIBLIOGRAPHY

8.1 Databases

Datastream. (2020) *Thomson Reuters Datastream*. [Online]. Available at: Subscription Service (Accessed: May 2020)

Organization for Economic Co-operation and Development (2020) 'Long-Term Government Bond Yields: 10-year: Main (Including Benchmark) for the Euro Area (IRLTLT01EZA156N)', *retrieved from Federal Reserve Bank of St. Louis*. Available at: https://fred.stlouisfed.org/series/IRLTLT01EZA156N (Accessed: May 2020)

8.2 Academic writings and articles

Balkema, G. and Embrechts, P. (2018) 'Linear regression for heavy tails', *Risks*, 6(3), pp. 1–70. doi: 10.3390/risks6030093.

Benabou, R. and Tirole, J. (2010) 'Individual and corporate social responsibility', *Economica*, 77(305), pp. 1–19. doi: 10.1111/j.1468-0335.2009.00843.x.

Chatterjee, S. *et al.* (2018) 'Fund Characteristics and Performances of Socially Responsible Mutual Funds: Do ESG Ratings Play a Role?', *Journal of Accounting and Finance*, 18(6). doi: 10.33423/jaf.v18i6.449.

Cheng, B., Loannou, L. and Serafeim, G. (2014) 'Corporate Social Responsibility And Access To Finance', *Strategic Management Journal*, 35(1), pp. 1–23. Available at: http://ssrn.com/abstract=1847085.

Christensen, H. B., Hail, L. and Leuz, C. (2019) 'Economic Analysis of Widespread Adoption of CSR and Sustainability Reporting Standards', *SSRN Electronic Journal*. doi: 10.2139/ssrn.3315673.

Cui, J., Jo, H. and Na, H. (2018) 'Does Corporate Social Responsibility Affect Information Asymmetry?', *Journal of Business Ethics*. Springer Netherlands, 148(3), pp. 549–572. doi: 10.1007/s10551-015-3003-8.

Dickinson, G. (1990) 'Geneva papers on risk and insurance. Issues and practice.', *The Geneva Papers on Risk and Insurance - Issues and Practice*, 26(3), pp. 360–366. doi: 10.1057/palgrave.gpp.25.

van Duuren, E., Plantinga, A. and Scholtens, B. (2016) 'ESG Integration and the Investment Management Process: Fundamental Investing Reinvented', *Journal of Business Ethics*. Springer Netherlands, 138(3), pp. 525–533. doi: 10.1007/s10551-015-2610-8.

Eccles, R. G., Kastrapeli, M. D. and 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), pp. 125–133. doi: 10.1111/jacf.12267.

Fama, E. F. and French, K. R. (1993) 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics*, 33(1), pp. 3–56. doi: 10.1016/0304-405X(93)90023-5.

Fama, E. F. and French, K. R. (1997) 'Industry costs of equity', *Journal of Financial Economics*, 43(2), pp. 153–193. doi: 10.1016/S0304-405X(96)00896-3.

Fama, E. and French, K. R. (2014) *A Five-Factor Asset Pricing Model*, *SSRN Electronic Journal*. doi: 10.1007/978-3-319-20612-7_39.

Ferrell, A., Liang, H. and Renneboog, L. (2016) 'Socially responsible firms', *Journal of Financial Economics*, 122(3), pp. 585–606. doi: 10.1016/j.jfineco.2015.12.003.

Friedman, M. (1970) 'The social responsibility of business is to increase its profits', *The New York Times Magazin*, 13 September, pp. 31–35.

Garriga, E. and Melé, D. (2004) 'Corporate social responsibility theories: Mapping the territory', *Journal of Business Ethics*, 53(1–2), pp. 51–71. doi: 10.1023/B:BUSI.0000039399.90587.34.

Gebhardt, W. R., Lee, C. M. C. and Swaminathan, B. (2001) 'Toward an implied cost of capital', *Journal of Accounting Research*, 39(1), pp. 135–176. doi: 10.1111/1475-679X.00007.

Goerg, G. M. (2011) 'Lambert W random variables-a new family of generalized skewed distributions with applications to risk estimation', *Annals of Applied Statistics*, 5(3), pp. 2197–2230. doi: 10.1214/11-AOAS457.

Goerg, G. M. (2015) 'The Lambert Way to Gaussianize Heavy-Tailed Data with the Inverse of Tukey's h Transformation as a Special Case', *Scientific World Journal*, 2015. doi: 10.1155/2015/909231.

Höllerer, M. A. (2012) 'Corporate social responsibility (CSR) Book Title : Between Creed , Rhetoric Façade , and Disregard Book Subtitle : Dissemination and Theorization of Corporate Social Responsibility in Austria', in *Between Creed, Rhetoric Façade, and Disregard: Dissemination and Theorization of Corporate Social Responsibility in Austria*. Available at: http://www.jstor.com/stable/j.ctv9hj7c3.7.

Kabašinskas, A. *et al.* (2009) 'Alpha-stable paradigm in financial markets', *Journal of Computational Analysis and Applications*, 11(4), pp. 641–668.

Kotsantonis, S. and Serafeim, G. (2019) 'Four Things No One Will Tell You About ESG Data', *Journal of Applied Corporate Finance*, 31(2), pp. 50–58. doi: 10.1111/jacf.12346.

Kring, S. *et al.* (2009) 'Estimation of α-Stable Sub-Gaussian Distributions.pdf', *Risk Assessment Decisions in Banking and FinanceDecisions in Banking and Finance*, pp. 111–152.

Manchiraju, H. and Rajgopal, S. (2017) 'Does Corporate Social Responsibility (CSR) Create Shareholder Value? Evidence from the Indian Companies Act 2013', *Journal of Accounting Research*, 55(5), pp. 1257–1300. doi: 10.1111/1475-679X.12174.

Masulis, R. W. and Reza, S. W. (2015) 'Agency problems of corporate philanthropy', *Review of Financial Studies*, 28(2), pp. 592–636. doi: 10.1093/rfs/hhu082.

Oorschot, J. A. (2017) *Tail dependence : OLS estimators under heavy-tails*. *Masters Thesis, Econometrie*, Erasmus University Rotterdam. Available at: http://hdl.handle.net/2105/39555.

Puaschunder, J. M. (2016) 'On the Emergence, Current State and Future Perspectives of Socially Responsible Investment (Sri)', *SSRN Electronic Journal*, 16(16), pp. 38–63. doi: 10.2139/ssrn.2720686.

Stoyanov, S., Samorodnitsky, G. and Rachey, S. (2007) 'Asymptotic distribution of unbiased linear estimators in the presence of heavy-tailed stochastic regressors and residuals', *Probability and Mathematical Statistics*. (Discussion Paper Series 1: Economic Studies No 21/2005), 27(21), pp. 275–302.

The Economist (2020) 'Companies are tying their loans to measures of do-goodery'.

Xu, G. and Genton, M. G. (2015) 'Efficient maximum approximated likelihood inference for Tukey's g-and-h distribution', *Computational Statistics and Data Analysis*, 91, pp. 78–91. doi: 10.1016/j.csda.2015.06.002.

Young In, S., Young Park, K. and Monk, A. (2018) 'Is 'Being Green' Rewarded in the Market? An Empirical Investigation of Decarbonization and Stock Returns.' *Stanford Global Projects Center Working Paper*. Available at: https://ssrn.com/abstract=3020304.

IX. ANNEX

9.1 Summary statistics

DAX Dataset

	Mean	Median	S.D.	Min	Max
RET	0.1464	0.1114	0.4333	-0.9105	5.960
RET_W	0.1335	0.1114	0.3510	-0.4709	0.8723
RET_G	0.1187	0.1114	0.3456	-0.8523	1.559
ESGSC	0.5095	0.4792	0.1588	0.1617	0.9333
ESGSCE	0.6519	0.6853	0.2142	0.1247	0.9916
ESGSCS	0.6661	0.6954	0.2034	0.06310	0.9906
ESGSCG	0.5189	0.5269	0.2090	0.03770	0.9828
HBETA	0.8700	0.7990	0.5702	-3.013	6.253
l_PROF	-2.294	-2.261	1.065	-8.126	2.107
MB	0.9067	0.4284	6.023	-27.50	240.2
gNREV	0.1951	0.06349	2.201	-0.9862	82.52

STOXX600 Dataset

	Mean	Median	S.D.	Min	Max
RET	0.1161	0.1027	0.3453	-0.9234	4.193
RET_W	0.1073	0.1027	0.2836	-0.4173	0.6560
RET_G	0.1008	0.1027	0.2837	-0.7314	1.251
ESGSC	0.5313	0.5047	0.1615	0.08410	0.9391
ESGSCE	0.6730	0.7004	0.1944	0.02630	0.9932
ESGSCS	0.6477	0.6755	0.1963	0.03620	0.9917
ESGSCG	0.5520	0.5682	0.2107	0.01800	0.9900
HBETA	0.9557	0.9070	0.8253	-27.16	49.70
l_PROF	-2.006	-1.997	1.021	-10.25	7.249
MB	1.011	0.5306	8.026	-130.1	691.6
gNREV	0.7367	0.05616	55.43	-15.40	5566

9.2 Units and reference codes for variables

Variable	Units	Description	Datastream Code
Symbol	—	Ticker Symbol	WC05601
P	€	Price	P
MV	€*1000	Market Value	MV
NOS	—	Number of Shares	NOSH
DPS	€	Dividend per Share	DPS
DY	Ratio	Dividend Yield	DY
HBETA	Ratio	Historical Beta	897E
ESGSC	Ratio	ESG Total Score	TRESGCS
ESGSCE	Ratio	ESG Environment Pillar Score	ENSCORE
ESGSCS	Ratio	ESG Social Pillar Score	SOSCORE
ESGSCG	Ratio	ESG Governance Pillar Score	CGSCORE
NREV	€*1000	Net Revenue	WC01001
EBIT	€*1000	Earnings before interest and taxes	WC18191
EBITDA	€*1000	EBIT & depreciation	WC18198
EPS	€	Earnings per share	EPS
PE	Ratio	Price/Earnings Ratio	PE
ROE	Ratio	Return on Equity	WC08301
ROA	Ratio	Return on Assets	WC08326
ROIC	Ratio	Return on invested capital	WC08376
GENINDCLASS	—	General industry classification	WC06010
COUNTRY	—	I/B/E/S country code	IBCTRY