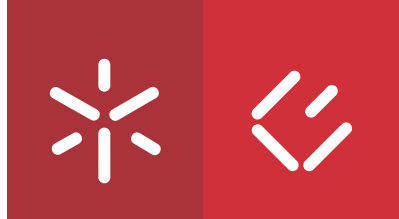


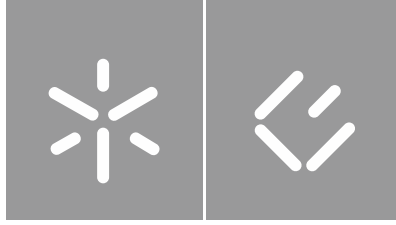


Sara Filipa Ribeiro Rodrigues

Carbon Emissions and Crash Risk:
US evidence from firm-level data

Universidade do Minho
Escola de Economia e Gestão





Universidade do Minho
Escola de Economia e Gestão

Sara Filipa Ribeiro Rodrigues

Carbon Emissions and Crash Risk: US evidence from firm-level data

Dissertação de Mestrado
Mestrado em Finanças
Trabalho efetuado sob a orientação do

Professor Doutor Nelson Manuel Pinho Brandão
Costa Areal

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição-NãoComercial-SemDerivações
CC BY-NC-ND

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Acknowledgments

First and foremost, I would like to thank my supervisor, Professor Nelson Areal, for bringing the weight of his considerable experience and knowledge to this dissertation and for his guidance throughout this journey.

A heartfelt thank you to my family, especially my parents and my brother, and close friends for their love and encouragement.

A special thanks to my boyfriend for his unwavering support and stimulating discussions. His presence has been a significant source of motivation.

Thank you all.

Statement of Integrity

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Resumo

Num mundo cada vez mais consciente das questões climáticas, compreender as implicações financeiras das emissões de carbono tornou-se crucial. À medida que investidores e governadores lidam com riscos associados a fatores ambientais, é vital examinar de que forma as emissões de carbono influenciam a estabilidade financeira e a dinâmica dos mercados. Esta dissertação investiga a relação entre as emissões de carbono e o risco de queda do preço das ações através de uma análise de 1,279 empresas listadas nos EUA, no período de 1999 a 2022. O estudo baseia-se no enquadramento teórico proposto por Jin and Myers (2006), que sugere que a assimetria de informação leva a colapsos nos preços das ações quando informações ocultas são reveladas. Partindo da premissa de que empresas com maiores emissões de carbono reportadas são mais transparentes e genuínas, o objetivo central é avaliar se estas empresas estão associadas a um menor risco de colapso. Duas medidas de risco de queda são utilizadas – Assimetria Condicional Negativa (*NCSKEW*) e Volatilidade (*DUVOL*) – para avaliar esta relação. Os resultados revelam uma associação não significativa entre as emissões de carbono e o risco de queda, medido pela *NCSKEW* e pela *DUVOL*, após controlar para os fatores que preveem o risco de queda e para os efeitos fixos de ano e empresa. Estudos futuros deveriam expandir a análise para uma amostra mais ampla e diferentes contextos geográficos de forma a retirar comparações e resultados mais robustos.

Palavras-chave: Assimetria Condicional Negativa, Assimetria de Informação, Emissões de Carbono, Risco de Queda do Preço das Ações, Volatilidade.

Abstract

In today's increasingly climate-conscious world, understanding the financial implications of carbon emissions has become crucial. As investors and policymakers grapple with the risks associated with environmental factors, it is vital to examine how carbon emissions influence financial stability and market dynamics. This dissertation investigates the relationship between carbon emissions and stock price crash risk through an analysis of 1,279 U.S. listed firms over the period from 1999 to 2022. The study is grounded in the theoretical framework proposed by Jin and Myers (2006), which suggests that information asymmetry leads to stock price crashes when hidden information is revealed. Based on the premise that companies with higher reported carbon emissions are more transparent and genuine, the main purpose is to evaluate whether these companies are associated with a lower risk of stock price collapse. Two measures of crash risk – Negative Conditional Skewness (*NCSKEW*) and Down-to-Up Volatility (*DUVOL*) – are employed to assess this relationship. The findings reveal no statistically significant association between carbon emissions and crash risk as measured by *NCSKEW* and *DUVOL*, after controlling for predictors factors of crash risk and year, and firm fixed effects. Further studies should extend the analysis to a broader sample and different geographical contexts to enhance comparative insights.

Keywords: Carbon Emissions, Down-to-Up Volatility, Information Asymmetry, Negative Conditional Skewness, Stock Price Crash Risk.

List of Contents

1.	Introduction	1
2.	Literature Review	3
3.	Methodology	5
3.1.	The Sample.....	5
3.2.	Carbon Emissions Measure	5
3.3.	Crash Risk Measures.....	5
3.4.	Control Variables	7
3.5.	Empirical Model	9
4.	Data	10
4.1.	Initial Stock Data	10
4.2.	Data Filters	10
4.2.1.	Stock filters based on static information	11
4.2.2.	Stock and stockday filters based on return index information	13
4.2.3.	Stock filters based on carbon emissions measure and industry classification	17
4.3.	Financial Accounting Variables.....	17
4.3.1.	Share Turnover.....	19
4.3.2.	Book Value of Equity.....	19
5.	Empirical Results	20
5.1.	Descriptive Statistics	20
5.2.	Pearson Correlation Matrix	23
5.3.	Findings	26
6.	Conclusions.....	28
	References.....	29
	Appendices.....	33
A.	Data Variables Description: Datastream Datatypes	33
B.	Data Variables Description: ESG Datatypes	35
C.	Data Variables Description: Worldscope Datatypes	36
D.	Industry Classification of Sample Firm	38

List of Tables

Table 1: Filters based on static information.....	12
Table 2: Filters based on return index information.	16
Table 3: Control variables description and sourced datatypes.....	18
Table 4: Mean values for crash risk and lagged carbon emissions measures (2000-2022).	21
Table 5: Summary of descriptive statistics.	23
Table 6: Pearson Correlation Matrix.....	25
Table 7: Carbon Emissions on Crash Risk - Regression Results.	27

List of Symbols

$r_{j,\tau}$	Return of stock j on day τ
$r_{m,\tau}$	Return on the value-weighted market portfolio m on day τ
$\varepsilon_{j,\tau}$	Error term of the extended market model
$R_{j,\tau}$	Firm-specific daily returns for stock j on day τ
n	Number of firm-specific daily returns for stock j during year t
n_d	Number of down days for stock j in year t
n_u	Number of up days for stock j in year t
$\sum R_{j,\tau}$	Sum of firm-specific daily returns for stock j during year t
$\sum_{Down} R_{j,\tau}$	Sum of firm-specific daily returns below the average for stock j in year t
$\sum_{Up} R_{j,\tau}$	Sum of firm-specific daily returns above the average for stock j in year t
$NCSKEW$	Negative Conditional Skewness
$DUVOL$	Down-to-Up Volatility
$CARBON$	Carbon Emissions (in tons of direct CO ₂ and CO ₂ equivalent) per US dollars of Revenues or Net Sales
$DTURN$	Detrended Turnover
RET	Past Returns
$SIZE$	Firm Size
$SIGMA$	Stock Volatility
LEV	Financial Leverage
ROA	Return on Assets or Profitability

1. Introduction

Carbon emissions are one of the primary drivers of climate change, which is causing a wide range of negative impacts on the planet. Governments and international organizations are increasingly focused on reducing carbon emissions and as such policymakers need to understand the effectiveness of different procedures and identify new approaches to limit the impacts of climate change. The Paris Agreement is a landmark international treaty adopted by the United Nations Framework Convention on Climate Change (UNFCCC) in December 2015. The Agreement sets out a global framework to limit global warming to well below 2 degrees Celsius above pre-industrial levels and to pursue efforts to limit the temperature increase even further (Delbeke et al., 2019). It builds on the previous international climate agreements, including the Kyoto Protocol and the Cancun Agreement, and it provides a strong foundation for continued progress towards a more sustainable and resilient future. In the United States (US), there have been several executive orders and actions taken by the Joe Biden administration aimed at combating climate change, including rejoining the Paris Agreement, a commitment made in response to the prior administration's decision by Donald Trump in June 2017, to withdraw from the accord. Additionally, the recent approval of the Inflation Reduction Act of 2022, recognized as the most significant clean energy and climate legislation in the US history, will provide significant investments in clean energy over the next 10 years, and can help reduce the nation's carbon emissions by roughly 40% by 2030, putting the US within reach of its international commitments (Bistline et al., 2023).

As part of these efforts, international organizations are progressively introducing policies aimed at pricing carbon emissions, either through carbon taxes or through cap-and-trade systems (Stavins, 2019). A consensus exists among economists and various policy analysts that, especially in the long-term, implementing economy-wide carbon pricing will be a crucial component of any effective policy aimed at achieving cost-effective reductions in carbon emissions in the US.

This shift in mindset is notably affecting financial markets. As the risks associated with climate change become more apparent, investors are increasingly looking to divest from fossil fuel investments, such as coal, oil, and gas, and to redirect their investments towards renewable energy sources. Monasterolo and de Angelis (2020) found that stock market investors have started to consider low-carbon assets as an appealing investment opportunity, especially after the Paris Agreement. In the same vein, Bolton and Kacperczyk, (2021) outlined the increased awareness of carbon risk among investors as it is an important consideration for their investment decision-making. On this context, it becomes more evident that companies that fail to address this subject may face increased scrutiny and may struggle to attract investments in the long term. Nevertheless, if it is true that investors are genuinely concerned about climate change and intend to reallocate their investments to address this issue, there will be a growing demand for transparent and accurate information. However, the lack of disclosure from firms still constitutes a barrier for investors to

better manage their portfolios' climate change exposure (Venturini, 2022). Additionally, there is the threat of greenwashing in relation to carbon emission reductions, an issue In and Schumacher (2021) dubbed "*carbonwashing*". This potentially leads to another challenge which is the information asymmetry between managers and market participants. According to previous studies, voluntary greenhouse gas emissions disclosure is associated with lower information asymmetry, indicating that it provides investors with more accurate and reliable information about a company's environmental performance (Borghei et al., 2018). Moreover, Adhikari and Zhou (2021) observed that US companies opting not to disclose carbon emission information or providing incomplete environmental data experience higher information asymmetry, as measured by bid-ask spreads. Generally, the market responds to first time carbon emission disclosure with decreases in the relative bid-ask spread and only companies that provide complete information and opt to make it available to the public enjoy the largest reduction. Overall, these studies suggest that voluntary carbon emissions can play a crucial role in mitigating information asymmetry.

This research aims to investigate whether companies with higher carbon emissions, being more transparent in their reporting, are less likely to experience a stock price crash. The study is based on the model developed by Jin and Myers (2006), which argues that stock price crashes result from information asymmetry. Given the urgent need for companies to achieve environmental goals and the significant impact of stock price crashes on investors' wealth, this research question remains highly relevant and worthy of exploration.

To develop the empirical study, data on US market-listed stocks between 1999 and 2022 will be used. The data is retrieved from Datastream and cleaned and processed according to Landis and Skouras (2021) guidelines.

The remainder of this Dissertation is structured as follows. Section 2 reviews the relevant literature for the current research; Section 3 presents the methodology, including the measurement of crash risk and carbon emissions as well as detailed explanation of the control variables used; Section 4 outlines the data required for analysis and the sample selection procedure; Section 5 presents and discusses the findings; while Section 6 offers concluding remarks for this Dissertation.

2. Literature Review

A large literature has tracked the impact of carbon emissions and other environmental-related subjects on several proxies of risk.

Kleimeier and Viehs (2018) revealed that higher carbon emissions, when adjusted for industry and firm size, contribute to increased loan spreads. Building on this, recent research by Caragnano et al. (2020) further expanded the understanding to the European capital markets by affirming that firms with higher carbon emissions intensity face substantially increased costs for debt financing. This underscores the notable impact of carbon emissions on financial operations. In the same vein, the findings of Li et al. (2014) indicated a positive correlation between a company's emission intensity and the cost of debt, while providing limited evidence for any effect on the cost of equity. The results from Trinks et al. (2022) uncovered a significant positive relationship between carbon intensity and the cost of equity capital, providing international evidence. This impact was particularly pronounced in sectors with high emissions and companies subject to carbon pricing regulation. These studies indicate that investors, both equity and debt holders, commonly seek a premium to mitigate the additional risk linked to a company's emissions level.

Another area of research examines the correlation between a company's carbon emissions and its distance-to-default, a prevalent market-based gauge of corporate default risk. Both Capasso et al. (2020) and Kabir et al. (2021) observed that higher emissions negatively impact a company's distance-to-default, suggesting that credit risk linked to a company's emissions is crucial for policy decisions made by relevant stakeholders. Furthermore, in their investigation, Aljughaiman et al. (2024) explored the repercussions on corporate tail risk and identified that companies with higher greenhouse gas emissions face a more severe penalty from shareholders. This is manifested in an elevated probability of experiencing extreme negative returns, commonly known as tail risk.

Building upon these insights, this research shifts its focus to examine another crucial risk proxy – stock price crash. This area has received noteworthy attention in recent finance literature, as evidenced by (Habib et al., 2018). For investors, the risk of a stock price crash stands out as a significant concern as it comes in irregular cycles, is typically unpredictable and can lead to severe losses. According to Jin and Myers, (2006), these crashes often result from an accumulation of bad news, and when it reaches a certain threshold, all this information comes at once leading to a sudden collapse in stock prices. The underlying premise here is rooted in information asymmetry between corporate insiders and external investors, with negative information not being disclosed in a timely manner but instead disclosed in a mass. This phenomenon aligns with the concept of conditional skewness, as outlined by Chen et al. (2001), representing the third moment of the return distribution. The occasional display of very large drops but not

equally large up moves is a well-known stylized fact of stock returns, meaning that the return distribution is asymmetric or negatively skewed (Chen et al., 2001; Christoffersen, 2012).

There is an extent field of studies aimed at capturing risk asymmetry and its determinants. Opaque financial reporting (Jin & Myers, 2006) and corporate tax avoidance (Kim et al., 2011b) are positively associated with firm-specific crash risk, while accounting conservatism is known to be a factor that can reduce the likelihood of a collapse (Kim & Zhang, 2016). Kim et al. (2011a) found that option incentives at the management level have a positive impact on crash risk, which is somewhat in line with the findings of Hunjra et al. (2020) that managerial ownership increases crash risk. This positive and significant relationship seems to be associated with agency problems.

In the field of Corporate Social Responsibility (CSR), Kim et al. (2014) found that this mechanism has a mitigating effect on stock price crash risk and this relation is more evident for firms with weak governance, which signals the active role of CSR. Consistent with this view, Hunjra et al. (2020) conclude that engagement in CSR strategies and an effective governance system have a negative impact on future stock price crashes for manufacturing firms in emerging economies. A more recent study focused on listed Chinese firms, remarks that companies more committed with environmental measures and targets significantly reduce the risk of a price crash (Liu et al., 2021). With slightly different results, a study by da Silva (2022) on the relationship between ESG disclosure and crash risk found no statistically significant association between these two variables in companies located in Emerging Markets and Developing Economies (EMDEs). Another interesting finding was that, in developed countries, governance disclosure is more effective in reducing crash risk than environmental and social disclosure. In the same line, the findings of Murata and Hamori (2021) suggested that ESG disclosures lowers future stock price crash risk, but the effect of ESG differ across the analysed regions (Europe, Japan, and the United States). Interestingly, in the case of the United States, ESG disclosure did not exhibit predictive power for stock price crash risk. Overall, while these views seem to signal the importance of pro-environmental strategies for financial stability, caution should be exercised in adopting a one-size-fits-all approach regarding CSR and ESG practices.

Extending the existing literature on carbon emissions and crash risk, the present research aims to investigate the potential risk mitigation associated with higher emitting companies in the context of a stock price crash, considering that these companies are more transparent and genuine in their reporting. The subsequent section outlines the empirical methodology, enabling a thorough examination of this relationship while accounting for various predictors of crash risk.

3. Methodology

3.1. The Sample

My sample comprises 1,279 listed firms in the United States spanning from December 1999 to December 2022, encompassing both actively traded stocks and stocks delisted during the analysis period. The sample and time-period selection will be explored further in Section 4. Appendix D displays the breakdown of sample firms by industry, revealing a predominant presence of companies in the Business Services industry (15.95%), with Electronic Equipment (6.96%), Retail (6.33%), and Petroleum and Natural Gas (6.10%) following closely behind.

3.2. Carbon Emissions Measure

To effectively measure carbon emissions, it is crucial to acquire a comprehensive understanding of the diverse categories of emissions a company may produce, along with methods to normalize this metric. This normalization is essential to ensure uniformity and comparability across the different companies within my sample.

According to the United States Environmental Protection Agency (EPA), scope 1 emissions comprise direct emissions from sources under the organization's control. Scope 2 includes indirect emissions linked to the acquisition of electricity, steam, heat, or cooling services. Conversely, scope 3 emissions extend beyond direct control or ownership, encompassing emissions from activities associated with assets not owned or controlled by the reporting entity but indirectly influenced within its value chain.

Building upon existing literature and acknowledging the limited availability of data in Datastream, this study will narrow its focus to scope 1 carbon emissions, recognized as an indispensable part of corporate carbon responsibility and management. These emissions will be scaled by the firm's net sales or revenues at the end of each year, providing insights into the carbon emissions (in tons of direct CO₂ and CO₂ equivalent) per US dollars of revenues or net sales for each firm (Aljughaiman et al., 2024; Qian & Schaltegger, 2017). The variable *CARBON* will serve this purpose in the empirical model, enabling to understand the relationship between the level of carbon emissions and the risk of a stock price crash.

3.3. Crash Risk Measures

Following the prior literature (Chen et al., 2001; Jin & Myers, 2006), two measures of firm-specific crash risk will be employed. These measures are based on firm-specific daily returns, $R_{j,\tau}$. This ensures that our crash risk measures indicate firm-specific factors instead of broad market movements (Kim et al., 2014). $R_{j,\tau}$ is computed as the natural logarithmic of one plus the residual return, $R_{j,\tau} = \ln(1 + \varepsilon_{j,\tau})$, of the following extended market model:

$$r_{j,\tau} = \alpha_j + \beta_{1,j}r_{m,\tau-2} + \beta_{2,j}r_{m,\tau-1} + \beta_{3,j}r_{m,\tau} + \beta_{4,j}r_{m,\tau+1} + \beta_{5,j}r_{m,\tau+2} + \varepsilon_{j,\tau} \quad (1)$$

where,

$r_{j,\tau}$ = return on stock j on day τ ,

$r_{m,\tau}$ = return on the value-weighted market portfolio m on day τ .

The return on stock j on day τ , $r_{j,\tau}$, is computed based on the daily changes in the Total Return Index (RI)¹ value retrieved from Datastream.

The return on the value-weighted market portfolio m on day τ , is sourced from the Kenneth R. French Library and provides the value-weighted return of all CRSP (Center for Research in Security Prices) firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ (Fama & French, 2015).

The lagged and leading market terms are included to allow for non-synchronous trading as different stocks have different trading frequencies (Dimson, 1979).

The residuals, $\varepsilon_{j,\tau}$, of the extended market model are not evenly distributed and are skewed. By computing the firm-specific daily returns, $R_{j,\tau}$, one can easily identify both crashes and positive jumps in the data symmetrically as the transformation helps to even out the distribution (Hutton et al., 2009).

The first measure of crash risk is the Negative Conditional Skewness, $NCSKEW$, which confines the asymmetry of the return distribution (Hunjra et al., 2020). It is calculated by taking the negative of the third moment of each stock's firm-specific daily returns for each year and normalizing it by the cubed standard deviation of firm-specific daily returns. Thus, for each firm j in year t ,

$$NCSKEW_{j,t} = -[n(n-1)^{3/2} \sum R_{j,\tau}^3] / [(n-1)(n-2)(\sum R_{j,\tau}^2)^{3/2}] \quad (2)$$

where,

$R_{j,\tau}$ = firm-specific daily return for firm j on day τ ,

n = number of daily returns during year t .

This measure is multiplied by -1 so that a higher value corresponds to a stock being more crash prone, i.e., having a more left-skewed distribution.

¹ In Appendix A, a detailed description of RI is given.

The second measure of crash risk is the Down-to-Up Volatility, *DUVOL*. For every stock j over a fiscal-year period t , firm-specific daily returns are separated into “down” days, when returns are below the period mean, and “up” days, when returns are above the period mean. The standard deviation is computed for each of these subsamples separately and then we take the log of the ratio of the standard deviation on the down days to the standard deviation on the up days. Specifically, *DUVOL* is calculated as follows:

$$DUVOL_{j,t} = \ln[(n_u - 1) \sum_{Down} R^2_{j,\tau} / (n_d - 1) \sum_{Up} R^2_{j,\tau}] \quad (3)$$

where,

n_u = number of up days in year t ,

n_d = number of down days in year t .

As suggested in Chen et al. (2001), this alternative measure does not involve third moments, and hence is less likely to be overly affected by extreme returns. Again, the convention is that a higher value of *DUVOL* indicates greater crash risk.

3.4. Control Variables

For this research, I decided to employ several control variables which researchers have previously found to have an influence on future crash risk. Chen et al. (2001) showed that trading volume, a proxy for the intensity of differences of opinion among investors, is a predictor of stock price crash risk. This is explained by the Hong and Stein (1999) model² which predicts that negative skewness in returns will be most pronounced around periods of heavy trading volume. Following their baseline specification, I used detrended turnover, *DTURN*, to account for this impact by computing the average monthly share turnover in year t minus the average monthly share turnover in year $t - 1$ (Kim et al., 2014; Yildiz & Karan, 2020). The reason for detrending is to remove any turnover component that can be considered as a relatively stable firm characteristic, adhering to a conservative approach. Detrending is then able to capture the intensity of disagreements in the market (Murata & Hamori, 2021).

Chen et al. (2001) discovered that historical returns and market-to-book ratios also play a role in predicting the risk of market crashes. These relationships are perhaps most clearly suggested by models of stochastic bubbles in which high past returns or high market-to-book ratios (glamour stocks) suggest a prolonged buildup of the bubble. Consequently, when the bubble bursts and prices revert to fundamental levels, the

² The Hong-Stein model relies on differences of opinion among investors regarding fundamental value and the presence of short-sales constraints for some. When bearish investors with constraints are forced to sell all shares due to differences of opinion, their information is not fully incorporated into prices, creating a corner solution. As more-bullish investors exit the market, the initially bearish group may become support buyers, revealing more about their signals. This process results in hidden information coming to light during market declines, leading to negatively skewed returns.

decline is more pronounced. This heightened decline is the underlying reason for both being anticipated to have a higher risk of crashing. I thus control for past returns, *RET*, calculated as the mean of firm-specific daily returns over the fiscal year times 100, and for the market-to-book ratio, *MB*, calculated as the market value of equity divided by the book value of equity (Kim et al., 2014).

Furthermore, it has been well documented in several studies (Chen et al., 2001; Harvey & Siddique, 2000) that skewness is more negative on average for large-cap firms. Specifically, Chen et al. (2001) developed a discretionary-disclosure hypothesis grounded in the notions that (i) managers tend to promptly reveal positive news but gradually release negative ones, and (ii) managers of small companies can hide bad news more easily using this approach. Hence, we control for firm size, *SIZE*, calculated as the natural logarithm of the market value of equity (Kim et al., 2014; Murata & Hamori, 2021).

The next control variable is stock volatility, *SIGMA*, as more volatile stocks are likely to be more crash prone. Foundational articles (Chen et al., 2001; Hutton et al., 2009) are based on a “volatility feedback” mechanism in which the release of good news, while positive, introduces market volatility, and bad news, in addition to its direct negative impact, is further magnified by an increased risk premium. However, some studies have found the opposite relationship between stock volatility and crash risk (Murata & Hamori, 2021). *SIGMA* is calculated as the standard deviation of firm-specific daily returns over the fiscal year.

In addition, I control for financial leverage, *LEV*, calculated as total long-term debt divided by total assets, and for profitability measured by return on assets, *ROA*, (Kim et al., 2014). Prior literature used leverage as a proxy for default risk but failed to find support for this proposition (Habib et al., 2018). For instance, Hutton et al. (2009) and Kim et al. (2011a, 2011b) find a negative association between leverage and crash risk, when leverage is expected to be positively correlated with bankruptcy risk (Campbell et al., 2008). According to Habib et al. (2018), one possible explanation for this unexpected finding is that highly leveraged firms might initially be undervalued by investors, reducing the likelihood of subsequent price crashes. As previously mentioned, glamour stocks are considered more crash-prone, therefore a positive relationship is expected between profitability, *ROA*, and crash risk.

3.5. Empirical Model

Following the empirical model developed by (Kim et al., 2014), in order to examine the impact of carbon emissions on future stock price crash risk, the following regression model will be estimated:

$$\begin{aligned} CRASH_RISK_{j,t} = & \beta_0 + \beta_1(CARBON_{j,t-1}) + \beta_2(DTURN_{j,t-1}) \\ & + \beta_3(RET_{j,t-1}) + \beta_4(MB_{j,t-1}) + \beta_5(SIZE_{j,t-1}) \\ & + \beta_6(SIGMA_{j,t-1}) + \beta_7(LEV_{j,t-1}) + \beta_8(ROA_{j,t-1}) \\ & + \beta_x(DYear) + \beta_y(DFirm) + \varepsilon_{j,t} \end{aligned} \quad (4)$$

In Equation 4, *CRASH_RISK* is the dependent variable and is a proxy for *NCSKEW* or *DUVOL*. *CARBON* is the main independent variable as it contains the carbon emissions measure. All independent variables are lagged by one year so that we can test if the carbon emissions level of firm *j* in year *t*−1, *CARBON*_{*j,t*−1}, predict crash risk in year *t*, *CRASH_RISK*_{*j,t*}.

The previously outlined control variables serve the purpose of accounting for firm-specific factors anticipated to influence the likelihood of future stock price crashes. Therefore, for each firm *j* in year *t*−1, we incorporate controls for changes in trading volume, *DTURN*_{*j,t*−1}; past returns, *RET*_{*j,t*−1}; market-to-book ratio, *MB*_{*j,t*−1}; firm size, *SIZE*_{*j,t*−1}; stock volatility, *SIGMA*_{*j,t*−1}; financial leverage, *LEV*_{*j,t*−1}; and return on assets, *ROA*_{*j,t*−1}, (Kim et al., 2014; Yildiz & Karan, 2020). All these firm-specific factors, including firm-specific daily returns, *R*_{*j,t*}, are winsorized at the 1st and 99th percentiles, a widely used technique in the literature to mitigate the impact of outliers (Zaman et al., 2021). The regression model also includes year and firm fixed effects to account for unobserved heterogeneity.

4. Data

The Refinitiv Eikon Datastream platform serves as the primary source for data in this dissertation. In subsection 4.1, detailed insights will be provided regarding the initial stock universe, while in subsection 4.2 I will present the subsequent application of filters to arrive at the final sample. Adhering to the guidelines outlined by Landis and Skouras (2021), particularly addressing static and return index information, was imperative to enhance data accuracy and mitigate survivorship bias. These guidelines, tailored for the US market, ensure a more nuanced approach to working with Datastream data, thereby contributing to the overall quality of the results. Finally, subsection 4.3 will delve into a comprehensive presentation of all firm-related information.

4.1. Initial Stock Data

To create the initial stock universe, while Landis and Skouras (2021) advocated for an alternative approach, which involved extracting information on all equities traded across every stock exchange within each country using TDS' Navigator GUI, I decided not to follow this method due to its time-consuming nature and limitations in data availability from the Datastream platform. Instead, I opted by relying on market-specific constituent lists provided by Datastream and Worldscope, an approach commonly employed by research teams working with data from these sources, as it is more efficient and aligns with established practices in the field (Karolyi et al., 2012; Schmidt et al., 2019).

Accordingly, different constituent lists were extracted and merged, including: (i) the FUSALL* lists, comprising all currently active stocks, (ii) the DEADUS* lists, consisting of delisted stocks, and (iii) the WSUS* lists provided by Worldscope. Worth mentioning that the stock universe should include not only stocks that were active at the time of the study, but also stocks that were inactive due to mergers, acquisitions, or failures, but show a portion of the price history over the entire analysis period (Landis & Skouras, 2021). Employing this method yielded a dataset comprising 116,834 stocks.

4.2. Data Filters

In this subsection, I will provide a comprehensive overview of the filtering process employed to address data challenges within the initial sample.

The filtering strategies encompass two main categories: stock filters, which involve excluding the entire historical record of an instrument, and stockday filters, which target specific days for individual stocks. These filters are subcategorized based on the type of information they use, including static data and return indexes. Considering the time-consuming nature of applying all filters provided by Landis and Skouras (2021), I opted to select those identified as most essential when working with Datastream data and most adaptable to my research. The authors provide users with the flexibility to apply their filters in the manner they deem

appropriate, emphasizing that their work aims to establish best practices for working with these data but acknowledging that there is no “source of truth”. This approach underscores the importance of exercising judgement in filter selection and implementation.

At the end of this subsection, I illustrate the need of removing specific stocks based on carbon emissions data availability and industry classification.

4.2.1. Stock filters based on static information

For each stock on each constituent list, the following variables were retrieved (Oliveira, 2023): Type of Instrument (TYPE), Datastream Code (DSCD), Base Date (BDATE), Expanded Security Name (ENAME), Datastream Exchange Mnemonic (EXMNEM), Geographical Classification (GEOGN), ISIN (ISIN), Primary Indicator Flag (ISINID), Code Local (LOC), Currency (PCUR) and Security Type Code (TRAC). In Appendix A, the description of these variables is given. Table 1 **Error! Reference source not found.** presents a description of the filters, that will be further explained, organized in the order in which they are applied in this Dissertation.

Firstly, I focused exclusively on instruments classified as equities, ensuring that all observations represented equity securities (where the datatype TYPE should be “EQ”). This filter is pertinent because constituent lists often include various types of securities, such as American Depositary Receipts (ADR), Closed-end Funds (CF), and Global Depositary Receipts (GDR), which could otherwise introduce noise or inaccuracies into the analysis.

Next, I narrowed down the dataset to stocks that are traded on US exchanges (EXMNEM). According to Landis and Skouras (2021), some researchers prefer to exclude stocks listed on secondary exchanges because they assume these stocks are likely to be small or traded over-the-counter (OTC). However, using this approach with Datastream data can lead to a problem as this database only provides information on the current exchange where a stock is listed. So, if researchers exclude secondary exchanges to avoid issues related to small stocks, they might also exclude stocks that got moved from main stock exchanges due to poor performance, which can create bias in the sample. Following Oliveira (2023), the US has 12 exchanges plus the OTC markets: NYSE (NYS), NYSE MKT (ASE), NYSE ARCA (XC), NASDAQ (NAS), NASDAQ/NMS (NMS), OTC Bulletin (XBQ), Non-Nasdaq OTC (OTC), Boston (BOS), Chicago (MID), Pacific (PSE), Philadelphia (PHL), and BARS (E1). Therefore, I will only consider stocks traded on these exchanges.

To refine the selection further, I included only instruments that could be classified as common stocks based on both the TRAC and ENAME variables. As noted by Landis and Skouras (2021), the classification of a stock as common remains consistent over time, thus using this filter helps prevent survivorship bias. Column 3 of Table 1 provides information of the valid security type codes and country-specific text strings that should

not be present in the extended names of stocks. Additionally, stocks with expanded names containing country-specific local listing identifiers were excluded, indicating potential listings in other countries. This additional filter aids in identifying cross-listed stocks and ensures their exclusion from the sample.

Moreover, I ensured that the dataset only consisted of companies domiciled in the United States (GEOGN) and securities traded in the US dollar (PCUR).

To prevent any duplicates from affecting my analysis, I removed any redundant entries based on both LOC and DSCD codes. The datatype LOC represents a code assigned to an instrument by the exchange where it is traded. When Datastream designates a stock as primary (datatype ISIND equals "P"), the authors noticed that all other stocks with the same local code are not common stocks. Therefore, it is recommended to exclude any stocks with a non-unique local code and whose ISINID is not "P", as long as there is at least one stock with this local code that does have ISINID = "P".

After applying these filters, the final sample consisted of 29,968 observations.

Table 1: Filters based on static information.

This table outlines all the filters applied to the static variables listed in the first column. It explains the rationale for each filter and specifies the accepted values used in the sample construction.

Variable	Purpose	Accepted Values
Type of Instrument (TYPE)	To guarantee that all observations are Equities.	'EQ'.
DataStream Exchange Mnemonic (EXMNEM)	To exclude stocks that are not traded on US exchanges.	'NAS', 'NYS', 'OTC', 'ASE', 'XSQ', 'XBQ', 'NMS', 'BOS', 'MID', 'PSE', 'PHL', 'E1'.
Security Type Code (TRAC)	To exclude stocks that cannot be classified as common stocks.	'ORD', 'ORDSUBR', 'FULLPAID', 'UNKNOWN', 'UNKNOW', 'KNOW', 'NA'.
Expanded Security Name (ENAME)	To exclude stocks identifiable as non-common stocks based on specific text strings within their extended names.	Do not contain: 'TRUST', 'REPR', 'RIGHT', 'SERIES', 'NV', 'IV', 'TST', 'REAL ESTATE INVESTMENT', 'REALTY', 'RLTY', 'ROYALTY INVESTMENT', 'ASSET INVESTMENT', 'CAPITAL INVESTMENT', 'ASSET MANAGEMENT', 'CAPITAL MANAGEMENT', 'INVESTMENT MANAGEMENT', 'VENTURE CAPITAL', 'FINANCIAL SHBI', 'PROPERTY INVESTORS', 'INCOME PROPERTY', 'UNITS', 'UNIT', 'LIMITED PARTENERSHIP', 'FUND', 'EQUITY PARTNERS', 'LIMITED VOTING', 'SUB VOTING', 'TIER ONE SUB', 'VARIABLE

		VOTING', 'NON VOTINGREIT', 'RESIDENTIAL', 'REIT', 'BENEFICIAL', 'BENEFICIARY', 'BENEFIT INTEREST', 'BEN INTEREST', 'SH BEN INT', 'WARRANT', 'WRTS', 'LP', 'LP INTEREST', 'LP UT', 'HOLDINGS LP', 'PARTENERS UNIT', 'PART INT', 'UNIT PARTENERSHIP', 'UNIT LIMITED', 'MORTGAGE', 'REAL ESTATE', 'CERTIFICATE', 'NO PAR VALUE', 'HOLDING UNIT', 'BACKED', 'ST MIN', 'CORTS', 'TORPS', 'TOPRS', 'SECURITIES TRUPS', 'QUIPS', 'STRATS HIGH YIELD', 'TOTAL RETURN', 'DIVERSIFIED HOLDINGS', '(SICAV)', 'DEPOSITARY', 'DEPOSITOR', 'RECEIPT', 'REP & SHARES', 'GLOBAL SHARES', 'ADR', 'GDR', 'EXPD.', 'EXPIRED', 'DUPLICATE', 'CONVERTIBLE', 'CNVRT.', 'CONVRT.', 'EXCH.', 'DEBANTURE', '(DEB)', 'NIL PAID', 'STRUCTURED ASSET', 'CALLABLE', 'FLOATING RATE', 'ADJUSTABLE', 'REDEEMABLE', 'PAIRED CTF', 'CONSOLIDATED', 'INSURED', 'CAPITAL SHARES', 'DEBT STRATEGIES', 'LIQUIDATING', 'LIQUID UNIT', 'L UNIT', 'LASD', 'ACQUISITION', 'CAP UNIT', 'INCOME UNIT', 'PREFERRED', '(NYS)', '(NAS)', '(ASE)', '(OT C)', '(XSQ)', '(XQB)'.
Geographical Classification of company (GEOGN)	To exclude stocks that are not domiciliated in the US.	'UNITED STATES'.
Currency (PCUR)	To remove all securities traded in a currency other than the US dollar.	'U\$'.
Code Local (LOC) and Primary Indicator Flag (ISINID)	To remove duplicated observations and ensure the exclusion of any non-common stocks from the sample.	No duplicated codes. If duplicates are found, remove them only if ISINID is not equal to 'P', provided that at least one stock with this LOC has ISINID equal to 'P'.
Datastream Code (DSCD)	To remove duplicated observations.	No duplicated codes.

4.2.2. Stock and stockday filters based on return index information

For the 29,968 stocks previously found, I extracted the Total Return Index (datatype RI) daily from December 1999 to December 2022. The choice of this timeframe is underpinned by two principal considerations.

Firstly, data reliability for US stocks significantly improves after December 1984, a point emphasized by Landis and Skouras (2021) for ensuring dataset robustness. However, due to limitations in Datastream and the emergence of carbon emissions disclosure from companies in the late 1990s and early 2000s, I opted to begin data extraction from December 1999. Secondly, to include the most recent and pertinent information, I extended the dataset to cover data up until the end of 2022.

For the data extraction process, I used the Datastream's DPL function, ensuring precision by retaining up to six decimal points, which is the maximum available number of decimal places in the database. It is important to note that Datastream does not provide direct returns but instead offers a return index that monitors the hypothetical value of an investment with reinvested cash flows, notably dividends. Consequently, even small values in this index can significantly affect return calculations due to rounding issues, emphasizing the importance of this step.

For certain stocks, the return index is unavailable for any date, and as is standard in the literature, these instruments were consequently excluded.

For the first filtering step, I obtained information regarding each stock's status (datatype ESTAT) and its corresponding delisting date (datatype TIME). When a company's stock is delisted, it means it is no longer actively traded on a public exchange. Initially, certain stocks were found to have been delisted prior to December 1999, and as a result, they were promptly excluded from the sample. For stocks that underwent delisting after this date, it is important to recognize that there can be a delay between the last date for which trading data is available and the official date when the stock is marked as delisted in the database. Consequently, many stocks exhibit constant return indexes toward the conclusion of their trading history even when the series have been truncated at their delisting dates. To address this concern, the first step involved the removal of data related to stocks showing constant values beyond their delisting dates. Additionally, a comparison was made between the date of the last non-zero return and the delisting date. If a difference of more than 10 days was identified, the tenth and subsequent daily observations were excluded; otherwise, the delisting date retrieved from Datastream was used as the end date of the series. Stocks with constant return indexes throughout the entire series were also excluded to ensure data reliability³.

In the second phase of the filtering process, I started by computing the daily percent change in the Total Return Index for each day τ , or daily return (%), within the time series, taking into account the individual stock's initial and end dates.

³ Please note that I have applied the condition of excluding stocks with constant return indexes to both dead and active companies in my sample. This ensured that subsequent filters were applied smoothly, avoiding any potential issues.

I excluded stocks exhibiting daily returns of zero in more than 95% of their sample as Landis and Skouras (2021) suggest that this pattern may indicate a significant portion of missing data for those stocks. Subsequently, an implausibility filter was implemented to eliminate stocks for which more than 98% of the non-zero daily returns are either non-negative or non-positive. The authors observed a disparity in Datastream's approach, where it seemingly derived its return index calculation from a dividend yield factor, implying a dividend payout for every available date. This stands in contrast with CRSP data, where this factor was recorded solely on May 3, 1985. This characteristic serves as a clear indicator of problematic data, and the impact is alleviated through the application of this filter.

We also applied a filter to address extreme returns, which are often associated with mistakes in handling stock split⁴ adjustments. Specifically, if the daily return on day τ exceeded 100% (or fell below -50%), and the return on day $\tau + 1$ was below -50% (or surpassed 100%), both days were eliminated.

The rationale behind the next filter is to focus on days where there is meaningful trading activity and returns, excluding days that are likely holidays or when markets are closed. This filter removed specific days where either non-missing or non-zero daily returns account for less than 0.5% of the total number of stocks available on that day. Landis and Skouras (2021) have confirmed that this filter correctly identified US holidays by comparing to CRSP data.

The third step in the filtering process focuses on managing volatility. Recognizing the significance of standard deviation as an indicator of volatility, I calculated the standard deviation individually for each stock and applied the following criteria: (i) High Volatility: stocks exhibiting a daily standard deviation exceeding 40% are excluded from the dataset. This is a common phenomenon in Datastream data, as noted by the authors, attributed to factors such as missing observations (leading to padded return indexes and consequently zero returns), substantial errors (such as inaccurate adjustments for corporate events), or extreme illiquidity resulting in infrequent price updates for specific stocks; (ii) Low Volatility: stocks with a daily standard deviation below 0.01bps are also removed from the sample. The primary aim of this filter is to capture scenarios where Datastream fills in missing prices by replacing them with the most recent available price, resulting in zero returns.

In the final stage of this process, I added the following two filters: (i) if 30 consecutive return indexes were identical, all subsequent observations were excluded until the next change to mitigate the impact of padded prices causing staleness; (ii) I required a minimum of 120 valid daily observations, unless the stock's initial available observation falls within the last 120 days of the sample period, to avoid stocks with an unusually brief historical record.

⁴ Corporate action in which a company increases the number of its outstanding shares while reducing the price per share. The goal is to make the stock more affordable to a broader range of investors.

Following the implementation of these filters, my sample size narrowed down to 15,262 companies. Table 2 provides a summary of these filters.

Table 2: Filters based on return index information.

This table outlines all the filters applied to return index information listed in the first column. It indicates whether each filter pertains to individual stocks or specific trading days and whether it applies to the entire sample or only a subset. Additionally, it explains the rationale behind each filter.

Filter	Type and Application	Purpose
Return index data availability	Stock filter applied to all companies.	Remove stocks for which return index data is not available for any date.
No longer traded stocks	Stockday filter applied to dead companies.	Remove observations beyond the last trading date.
Few zero observations	Stock filter applied to all companies.	Remove stocks for which the daily return is zero on more than 95% of the days between their appearance in the database and the last date of the sample or the delisting date.
Implausibility	Stock filter applied to all companies.	Remove stocks of which more than 98% of their non-zero daily returns are either non-negative or non-positive.
Outlier errors	Stockday filter applied to all companies.	Remove observations if the daily return on day t is $> 100\%$ (or $< -50\%$) and the daily return on day $t + 1$ is $< -50\%$ (or $> 100\%$).
Holidays	Stockday filter applied to all companies.	Remove all observations in days on which non-missing or non-zero daily returns account for less than 0.5% of the total number of stocks available for that day.
High Volatility	Stock filter applied to all companies.	Remove stocks with a daily standard deviation of more than 40%.
Low Volatility	Stock filter applied to all companies.	Remove stocks with a daily standard deviation of less than 0.01bps.

Staleness	Stockday filter applied to all companies.	Remove observations when 30 consecutive identical return indexes are identified.
Few observations	Stock filter applied to all companies.	Remove stocks with less than 120 historical observations.

4.2.3. Stock filters based on carbon emissions measure and industry classification

The next phase involved gathering carbon emissions data for the remaining 15,262 companies in the sample. As detailed in section 0, this study specifically focuses on scope 1 emissions. Therefore, I obtained the CO2 Equivalent Emissions Direct, Scope 1 data (datatype ENERDP024)⁵ spanning from 1999 to 2022, with yearly observations. Upon analysis, it was found that only 1,679 companies had available information. Additionally, to construct the carbon emissions measure, data on Net Sales or Revenues (datatype WC01001)⁶ for the same timeframe was collected. However, 108 companies lacked the necessary Worldscope information, resulting in a reduction of the sample size to 1,571 companies.

Additionally, I collected information concerning the SIC (Standard Industrial Classification) codes for each company and merged it with the Fama-French 48 industry categories. This approach is extensively used in academic research, supported by numerous citations and an extensive body of literature dedicated to this classification (Phillips & Ormsby, 2016). Building on previous studies, I have excluded firms operating within the financial (SIC codes 6000-6999) and utilities (SIC codes 4900-4949) sectors from the analysis. This decision stems from the distinct accounting practices of financial firms, driven by their unique nature and regulatory requirements, as well as the stringent transparency regulations that utilities must adhere to (Hutton et al., 2009; Kim et al., 2014). Moreover, as highlighted by da Silva (2022), companies within these sectors often carry the perception of being “too big to fail”, accompanied by an implicit government bailout guarantee. As a result, 292 companies were excluded from the sample.

4.3. Financial Accounting Variables

For the 1,279 stocks in the sample, I collected financial accounting data needed for computing the explanatory variables outlined in Subsection 3.4 and described in Table 3.

⁵ In Appendix B, a detailed description of ESG datatypes is given.

⁶ In Appendix C, a detailed description of Worldscope datatypes is given.

Table 3: Control variables description and sourced datatypes.

This table outlines all the variables of control used in this empirical model, listed in the first column. It provides detailed descriptions of how each variable is computed and the sourced datatypes used for these calculations. All datatypes are thoroughly described in Appendix C.

Variables of Control	Description	Datatypes sourced from Datastream
Trading Volume	Average monthly share turnover in year t minus the average monthly share turnover in year $t - 1$.	- Trading Volume (WC08006) - Common Shares Outstanding (WC05301)
Past Returns	Mean of firm-specific daily returns, $R_{j,\tau}$, over the fiscal year.	- Total Return Index (RI)
Market-to-book ratio	Market value of equity divided by the book value of equity.	- Market Value (Capital) (MV) - Total Assets (WC02999) - Total Liabilities (WC03351) - Deferred Income Taxes & Investment Tax Credit (WC04101) - Preferred Stock (WC03451)
Firm Size	Natural logarithm of the market value of equity.	- Market Value (Capital) (MV)
Volatility	Standard deviation of firm-specific daily returns, $R_{j,\tau}$, over the fiscal year.	- Total Return Index (TRI)
Financial Leverage	Total long-term debt divided by total assets.	- Long-Term Debt (WC03251) - Total Assets (WC02999)
Profitability	Return on Assets.	- Return on Assets (WC08326)

4.3.1. Share Turnover

As indicated in Table 3, share turnover serves as the basis for calculating trading volume. According to previous literature, monthly share turnover is defined as the monthly share trading volume (WC08006) divided by the total number of shares outstanding (WC05301) for that month (Feng et al., 2022; Kim et al., 2014). Thus:

$$Share\ Turnover_{j,t} = \frac{Trading\ Volume_{j,t}(WC08006)}{Common\ Shares\ Outstanding_{j,t}(WC05301)} \quad (5)$$

4.3.2. Book Value of Equity

To derive the Book Value of Equity, equation 6 is employed, consistent with methodologies outlined by Fama and French (2006) and Oliveira (2023). For cases where Deferred Income Taxes & Investment Tax Credit and Preferred Stock are unavailable, they are presumed to be zero. Thus:

$$\begin{aligned} Book\ Value\ of\ Equity_{j,t} &= Total\ Assets_{j,t}(WC02999) - Total\ Liabilities_{j,t}(WC03351) \\ &+ Deferred\ Income\ Taxes\ \&\ Investment\ Tax\ Credit_{j,t}(WC04101) \\ &- Preferred\ Stock_{j,t}(WC03451) \end{aligned} \quad (6)$$

5. Empirical Results

5.1. Descriptive Statistics

Table 4 presents the mean values of stock price crash risk, $NCSKEW_t$ and $DUVOL_t$, and the lagged carbon emissions measure, $CARBON_{t-1}$, by year. In the early years, data on carbon emissions are entirely missing, indicating either a lack of reporting or availability issues.

Both $NCSKEW_t$ and $DUVOL_t$ exhibit moderate variation over the years, but they do not consistently follow the same pattern. During the financial crisis in 2008, there is a notable increase in crash risk, as indicated by both measures ($NCSKEW_t$ at 0.0042 and $DUVOL_t$ at 0.0016), followed by a decrease in 2009, which is way more pronounced for the $NCSKEW_t$ (-0.2204). This observation aligns with previous studies (Kim et al., 2014), which also identified heightened crash risk during the year 2008.

In 2018, there is also an increase in crash risk, more pronounced in the case of the $NCSKEW_t$ (0.1318). One possible explanation could be the election of Donald Trump as President of the United States on November 8, 2016, with his term beginning on January 20, 2017. Presidential elections tend to make financial markets highly volatile, particularly when radical policy changes are expected. In the case of 2017, these changes included a shift away from free trade and various “Brexit-type” actions involving entities like the UN (United Nations), UNESCO (United Nations Educational, Scientific, and Cultural Organization), the Paris Climate Accord, NAFTA (North America Free Trade Agreement), and TPP (Trans-Pacific Partnership) (Pham et al., 2018). Li et al. (2018) found that political events, specifically the elections of national leaders, increase political uncertainty, which creates incentives for managers to temporarily restrict the release of negative information about their firms. Consequently, firm stock is less likely to crash during the election years and more likely to crash in the post-election year, as the release of adverse news occur when the uncertainty is reduced, which might explain the peak observed in 2018.

Interestingly, in 2020, there is a considerable divergence in crash risk averages, with $DUVOL_t$ (0.0018) indicating higher risk of crash and $NCSKEW_t$ (-0.0719) lower risk. Considering that the global Covid-19 pandemic led to a significant drop in stock prices and increased market volatility (Dai et al., 2021), one would expect a higher crash risk during this period, as reflected in the $DUVOL_t$ value. However, it is worth noting that the market began to rebound around April 2020 with the Federal Reserve’s intervention, despite widespread economic shutdowns and rising unemployment. By the end of the year, amidst additional uncertainty from a contentious presidential election, the stock market still grew, suggesting that the $NCSKEW_t$ average may not be entirely random.

In 2022, both crash risk averages are positive, however $DUVOL_t$ (0.0006) remained quite stable when compared to the previous year, which is not the case for $NCSKEW_t$ (0.0395). It might also be related to

the negative impact of the Russian-Ukraine war on the global financial market which began in February 2022 (Assaf et al., 2023).

As for the $CARBON_{t-1}$, it shows relatively stable values over the years, indicating a gradual increase in reporting and potentially in emissions themselves, especially after 2015. This can be linked to the reporting obligations established by the Paris Agreement in December 2015. Under its Enhanced Transparency Framework, all Parties are mandated to report their greenhouse gas (GHG) balance biannually and to monitor the progress of individual countries in achieving their mitigation targets (Perugini et al., 2021). Nevertheless, a report developed by As You Sow⁷ in 2022, evaluated the effort of 55 major US corporations in reducing GHG emissions in alignment with the Paris Agreement's target of limiting global temperature rise to 1.5 degrees Celsius, aiming for net zero emissions by 2050. They found that most companies still lack comprehensive 1.5 degree-aligned GHG reduction goals and are not making sufficient progress toward net zero emissions. Moreover, even though most companies report Scope 1 emissions, disclosure of carbon offsets is often unclear, which underscores the critical need of exercising caution while looking at carbon emissions figures over time. Moreover, there does not appear to be any specific pattern between the lagged carbon emissions and crash risk variables.

Table 4: Mean values for crash risk and lagged carbon emissions measures (2000-2022).

This table presents the average values of crash risk, $NCSKEW_t$ and $DUVOL_t$, and lagged carbon emissions, $CARBON_{t-1}$, measures from 2000 to 2022. Refer to Section 3 for variables definition.

$YEAR_t$	$NCSKEW_t$	$DUVOL_t$	$CARBON_{t-1}$
2000	-0.2261	0.0028	-
2001	-0.1430	0.0018	-
2002	-0.0125	0.0016	-
2003	-0.1627	0.0007	0.2062
2004	-0.1923	0.0005	0.1496
2005	-0.1801	0.0004	0.1608
2006	-0.1777	0.0003	0.1357
2007	-0.1932	0.0005	0.1240
2008	0.0042	0.0016	0.1294
2009	-0.2204	0.0013	0.1417
2010	-0.1881	0.0004	0.1777

⁷ As You Sow is a non-profit organization that promotes environmental and social corporate responsibility through shareholder advocacy, coalition building, and innovative legal strategies. It engages in a variety of initiatives aimed at encouraging companies to adopt more sustainable and socially responsible practices. These initiatives include advocating for corporate transparency, reducing GHG emissions, promoting fair labor practices, and minimizing environmental impacts.

2011	-0.1036	0.0004	0.1467
2012	-0.0989	0.0003	0.1289
2013	-0.1080	0.0002	0.1291
2014	-0.0657	0.0003	0.1427
2015	-0.0042	0.0004	0.1426
2016	-0.0332	0.0007	0.1614
2017	0.0978	0.0003	0.1786
2018	0.1318	0.0006	0.1912
2019	0.0836	0.0004	0.1915
2020	-0.0719	0.0018	0.1804
2021	-0.0772	0.0007	0.2048
2022	0.0395	0.0006	0.1601

Table 5 presents the descriptive statistics of the sample. Building on the insights provided earlier, the mean values of crash risk measures, $NCSKEW_t$ and $DUVOL_t$, are -0.0724 and 0.0008, respectively. These values suggest that, on average, the sample firms have more right-skewed firm-specific daily returns and slightly higher volatility in firm-specific returns on down days compared to up days, although the difference is small. The medians (-0.0557 for $NCSKEW_t$ and 0.0001 for $DUVOL_t$) are slightly lower than the respective means, indicating the presence of a few relatively high values that raise the mean, while most data points are clustered towards the lower end of the range. For $NCSKEW_t$, the estimates are similar to those found by Chen et al. (2001) and Yildiz and Karan (2020) but differ from Kim et al. (2014) and Murata and Hamori (2021). For $DUVOL_t$, the mean is very close to zero, which is lower than the average reported in the existing literature, likely due to differences in sample periods and sample construction.

The average value of the total carbon emission per US dollars of revenues or net sales is 0.1682. The average change in monthly trading volume (as a percentage of shares outstanding) is 0.0214. Furthermore, the average firm in the sample has a firm-specific daily return of 13.27%, a market capitalization of \$3.24 billion⁸, a market-to-book ratio of 0.1650, a daily return volatility of 0.0209, a leverage of 0.2643, and a return on assets of 0.0490. RET_{t-1} and $SIZE_{t-1}$ have the highest standard deviations, suggesting (i) high variability in firm-specific daily returns, and (ii) diverse set of firms in the sample from small-cap to large-cap companies.

⁸ To interpret the mean value of the $SIZE$ variable in terms of actual market value, one must reverse the log transformation by exponentiating the mean value.

Table 5: Summary of descriptive statistics.

This table presents the descriptive statistics for the variables used in the dissertation. The data comprises a sample of 1,279 US-listed stocks from 1999 to 2022, including both active stocks and those that have been delisted but retain a portion of their price history throughout the analysis period. Financials (SIC codes 6000-6999) and utilities (SIC codes 4900-4949) are excluded, based on the Fama-French 48 industry classifications. Refer to Section 3 for variable definitions.

VARIABLES	MEAN	MEDIAN	STANDARD DEVIATION	25TH PERCENTILE	75TH PERCENTILE
<i>DUVOL_t</i>	0.0008	0.0001	0.0030	0.0000	0.0002
<i>NCSKEW_t</i>	-0.0724	-0.0557	0.5432	-0.3805	0.2596
<i>CARBON_{t-1}</i>	0.1682	0.0132	0.4194	0.0026	0.0848
<i>DTURN_{t-1}</i>	0.0214	0.0077	0.5948	-0.2087	0.2344
<i>RET_{t-1}</i>	0.1327	-0.0211	0.8778	-0.0956	0.0543
<i>MB_{t-1}</i>	0.1650	0.1137	0.1662	0.0629	0.2030
<i>SIZE_{t-1}</i>	8.0842	8.0577	1.7676	6.9209	9.2774
<i>SIGMA_{t-1}</i>	0.0209	0.0173	0.0128	0.0124	0.0250
<i>LEV_{t-1}</i>	0.2643	0.2376	0.1855	0.1323	0.3646
<i>ROA_{t-1}</i>	0.0490	0.0636	0.1199	0.0248	0.1035

5.2. Pearson Correlation Matrix

Table 6 reports the results of the Pearson correlation matrix. The highest correlations are observed between *DUVOL* and *SIGMA* (0.73), and *DUVOL* and *RET* (0.58), indicating that stocks with greater volatility and higher past returns are positively correlated with increased crash risk, as measured by *DUVOL*. Conversely, there is a negative correlation between *NCSKEW* and *SIGMA* (-0.22), and *NCSKEW* and *RET* (-0.43) within my sample. Additionally, there is a strong negative correlation between *SIGMA* and both *SIZE* (-0.53) and *ROA* (-0.42), indicating that smaller firms and those with lower profitability tend to experience higher volatility in firm-specific daily returns.

Furthermore, the positive and statistically significant correlation between *CARBON* and *DUVOL* indicates a weak positive relationship (0.1356), suggesting that firms with higher carbon emissions are associated with slightly higher down-to-up volatility. Conversely, the negative and statistically significant correlation between *CARBON* and *NCSKEW* reveals a very weak negative relationship, in contrast to the positive relationship found with *DUVOL*, indicating that higher carbon emissions might be associated with slightly

less negative skewness in stock returns. However, given the weak nature of these relationships, drawing robust conclusions is challenging.

5.3. Findings

Table 7 presents the OLS estimation results, with $NCSKEW_t$ as the dependent variable in Panel A, and $DUVOL_t$ as the dependent variable in Panel B. Both regressions include fixed effects at the year and firm levels.

Examining the results presented in Panel A, the model shows a negative association between the level of carbon emissions per US dollars of revenues or net sales, $CARBON_{t-1}$, and the risk of a stock price crash in the upcoming year, measured by $NCSKEW_t$. This relationship seems to be aligned with the initial expectation that being more genuine and transparent in reporting would associate with lower crash risk. However, this relationship did not achieve statistical significance, meaning that there is insufficient evidence to conclude a meaningful association between these variables within the context of the study. The only variable found to be statistically significant at the 1% level is $SIZE_{t-1}$ with a coefficient of 0.3091. This positive and significant coefficient suggests that larger firms are more likely to experience higher $NCSKEW_t$, indicating an increased risk of a stock price crash. This finding supports the discretionary-disclosure hypothesis proposed by Chen et al. (2001), which posits that large-cap firms tend to exhibit more negative skewness on average. This relationship is further corroborated by recent studies (Hunjra et al., 2020; Kim et al., 2014).

Turning to the results in Panel B, we observe a positive association between the level of carbon emissions per US dollars of revenues or net sales, $CARBON_{t-1}$, and the risk of a stock price crash in the upcoming year, $DUVOL_t$. However, this association also lacks statistical significance. Despite this, the positive relationship appears counterintuitive to my analysis. This discrepancy suggests that $DUVOL_t$ might be capturing a different dimension of carbon emissions over time. This indicates a need to investigate whether the relationship between emissions and crash risk is more complex than initially assumed. Additionally, the model identifies RET_{t-1} (0.0009***) and MB_{t-1} (0.0009*) as predictors of one-year ahead crash risk, with both showing positive and statistically significant relationships at the 1% and 10% levels, respectively. These associations seem to be aligned with the theory behind models of stochastic bubbles in which high past returns and high market-to-book ratios (glamour stocks) are associated with higher risk of stock price crash (Chen et al., 2001). Volatility, $SIGMA_{t-1}$ (0.1475***) and financial leverage, LEV_{t-1} , (-0.0005*) are also found to have a statistically significant impact on the risk of a stock price crash. These results are also in line with the literature, which suggests that high volatility and lower leveraged firms are linked to increased crash risk (Chen et al., 2001; Hutton et al., 2009; Kim et al., 2011a, 2011b).

Table 7: Carbon Emissions on Crash Risk - Regression Results.

This table presents the regression results of the effect of carbon emissions on firm-level stock price crash risk. The standard errors clustered at the firm and year levels are reported in parentheses. Refer to Section 3 for variable definitions. Note: *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

DEPENDENT VAR. =	<i>NCSKEW_t</i> PANEL A	<i>DUVOL_t</i> PANEL B
<i>CARBON_{t-1}</i>	-0.0633 (0.0606)	0.0005 (0.0003)
<i>DTURN_{t-1}</i>	-0.0254 (0.0191)	0.0001 (0.0001)
<i>RET_{t-1}</i>	0.0226 (0.0019)	0.0009*** (0.0003)
<i>MB_{t-1}</i>	0.2876 (0.1841)	0.0009* (0.0005)
<i>SIZE_{t-1}</i>	0.3091*** (0.0416)	-0.0001 (0.0001)
<i>SIGMA_{t-1}</i>	-1.9829 (2.4056)	0.1475*** (0.0308)
<i>LEV_{t-1}</i>	0.0126 (0.1038)	-0.0005* (0.0003)
<i>ROA_{t-1}</i>	-0.1247 (0.1881)	-0.0001 (0.0007)
CONSTANT	-0.0701 (0.0961)	-0.0021** (0.0006)
YEAR FE	Yes	Yes
COMPANY FE	Yes	Yes
OBSERVATIONS	5581	5581
R-SQUARED	0.302	0.751

6. Conclusions

In this Dissertation, the aim was to examine the effects of carbon emissions on the one-year ahead crash risk for a sample of 1,279 US-listed firms. The research was based on the model developed by Jin and Myers (2006), which argues that stock price crashes result from information asymmetry. When a certain threshold is reached, all the hidden information comes simultaneously, leading to the collapse in stock prices. Two measures of crash risk were employed: the Negative Conditional Skewness, $NCSKEW$, which captures the asymmetry of the return distribution, and the Down-to-Up Volatility, $DUVOL$, which assesses the asymmetry in stock return volatility between periods of positive and negative returns relative to a stock's average return.

The findings indicate no statistically significant relationship between carbon emissions, $CARBON_{t-1}$, and one-year ahead crash risk measured by both $NCSKEW_t$ and $DUVOL_t$. Regarding the variables of control, despite some findings lacking statistical significance, the analysis identified statistically significant relationships between crash risk and past returns, RET_{t-1} , volatility, $SIGMA_{t-1}$, firm size, $SIZE_{t-1}$, market-to-book ratio, MB_{t-1} , and financial leverage, LEV_{t-1} , in line with the literature.

It is important to acknowledge that this study encountered several data availability challenges, particularly regarding carbon emissions. Following the methodology of Landis and Skouras (2021), significant time and effort were required to compile an accurate sample. Due to these constraints, direct comparisons with existing literature may not be straightforward. Most prior studies used different databases that provided easier access to more comprehensive datasets, which were not available for this analysis. These limitations underscore the need for improved data accessibility and consistency in future research to facilitate more robust comparisons and conclusions.

For future research, it would be valuable to conduct further investigation into the various measures used to compute crash risk and explore how they might capture different information during the same event or time period. Expanding the analysis to a larger sample of firms and including companies from diverse countries could provide more meaningful comparisons and a richer understanding of the relationship between crash risk and carbon emissions in different contexts. Additionally, exploring whether companies with higher carbon emissions experience lower information asymmetry compared to those with lower emissions could offer valuable insights into how this characteristic influence crash risk dynamics. This further investigation may reveal nuances in how carbon emissions impact information transparency and, consequently, stock price stability.

References

- Adhikari, A., & Zhou, H. (2021). Voluntary disclosure and information asymmetry: do investors in US capital markets care about carbon emission? *Sustainability Accounting, Management and Policy Journal*, 13(1), 195-220.
- Aljughaiman, A. A., Cao, N. D., & Albarrak, M. S. (2024). The impact of greenhouse gas emission on corporate's tail risk. *Journal of Sustainable Finance & Investment*, 14 (1), 68-85.
- Assaf, R., Gupta, D., & Kumar, R. (2023). The price of war: Effect of the Russia-Ukraine war on the global financial market. *The Journal of Economic Asymmetries*, 28, e00328.
- Bistline, J., Blanford, G., Brown, M., Burtraw, D., Domeshek, M., Farbes, J.,...Jones, R. (2023). Emissions and energy impacts of the Inflation Reduction Act. *Science*, 380(6652), 1324-1327.
- Bolton, P., & Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of financial economics*, 142(2), 517-549.
- Borghei, Z., Leung, P., & Guthrie, J. (2018). Does voluntary greenhouse gas emissions disclosure reduce information asymmetry? Australian evidence. *Afro-Asian Journal of Finance and Accounting*, 8(2), 123-147.
- Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *The Journal of finance*, 63(6), 2899-2939.
- Capasso, G., Gianfrate, G., & Spinelli, M. (2020). Climate change and credit risk. *Journal of Cleaner Production*, 266, 121634.
- Caragnano, A., Mariani, M., Pizzutilo, F., & Zito, M. (2020). Is it worth reducing GHG emissions? Exploring the effect on the cost of debt financing. *Journal of Environmental Management*, 270, 110860.
- Chen, J., Hong, H., & Stein, J. C. (2001). Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices. *Journal of Financial Economics*, 61(3), 345-345-381.
- Christoffersen, P. (2012). *Elements of financial risk management* (2nd ed.). Academic press.
- da Silva, P. P. (2022). Crash risk and ESG disclosure. *Borsa Istanbul Review*, 22(4), 794-811.
- Dai, P.-F., Xiong, X., Liu, Z., Huynh, T. L. D., & Sun, J. (2021). Preventing crash in stock market: The role of economic policy uncertainty during COVID-19. *Financial Innovation*, 7, 1-15.

- Delbeke, J., Runge-Metzger, A., Slingenberg, Y., & Werksman, J. (2019). The paris agreement. In *Towards a climate-neutral Europe* (pp. 24-45). Routledge.
- Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of financial economics*, 7(2), 197-226.
- Fama, E. F., & French, K. R. (2006). Profitability, investment and average returns. *Journal of financial economics*, 82(3), 491-518.
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, 116(1), 1-22.
- Feng, J., Goodell, J. W., & Shen, D. (2022). ESG rating and stock price crash risk: Evidence from China. *Finance Research Letters*, 46, 102476.
- Habib, A., Hasan, M. M., & Jiang, H. (2018). Stock price crash risk: review of the empirical literature. *Accounting & Finance*, 58, 211-251.
- Harvey, C. R., & Siddique, A. (2000). Conditional skewness in asset pricing tests. *The Journal of finance*, 55(3), 1263-1295.
- Hong, H., & Stein, J. C. (1999). *Differences of opinion, rational arbitrage and market crashes*. National bureau of economic research Cambridge, Mass., USA.
- Hunjra, A. I., Mehmood, R., & Tayachi, T. (2020). How do corporate social responsibility and corporate governance affect stock price crash risk? *Journal of Risk and Financial Management*, 13(2), 30.
- Hutton, A. P., Marcus, A. J., & Tehranian, H. (2009). Opaque financial reports, R2, and crash risk. *Journal of financial Economics*, 94(1), 67-86.
- In, S. Y., & Schumacher, K. (2021). Carbonwashing: A New Type of Carbon Data-Related ESG Greenwashing. *Available at SSRN*.
- Jin, L., & Myers, S. C. (2006). R2 around the world: New theory and new tests. *Journal of Financial Economics*, 79(2), 257-257-292.
- Kabir, M. N., Rahman, S., Rahman, M. A., & Anwar, M. (2021). Carbon emissions and default risk: International evidence from firm-level data. *Economic Modelling*, 103, 105617.

- Karolyi, G. A., Lee, K.-H., & Van Dijk, M. A. (2012). Understanding commonality in liquidity around the world. *Journal of financial economics*, 105(1), 82-112.
- Kim, J.-B., Li, Y., & Zhang, L. (2011a). CFOs versus CEOs: Equity incentives and crashes. *Journal of Financial Economics*, 101(3), 713-713-730.
- Kim, J.-B., Li, Y., & Zhang, L. (2011b). Corporate tax avoidance and stock price crash risk: Firm-level analysis. *Journal of Financial Economics*, 100(3), 639-639-662.
- Kim, J. B., & Zhang, L. (2016). Accounting Conservatism and Stock Price Crash Risk: Firm-level Evidence. *Contemporary Accounting Research*, 33(1), 412-412 - 441.
- Kim, Y., Li, H., & Li, S. (2014). Corporate social responsibility and stock price crash risk. *Journal of Banking & Finance*, 43, 1-13.
- Kleimeier, S., & Viehs, M. (2018). Carbon disclosure, emission levels, and the cost of debt. *Emission Levels, and the Cost of Debt (January 7, 2018)*.
- Landis, C., & Skouras, S. (2021). Guidelines for asset pricing research using international equity data from Thomson Reuters Datastream. *Journal of Banking & Finance*, 130, 106128.
- Li, Q., Li, S., & Xu, L. (2018). National elections and tail risk: International evidence. *Journal of Banking & Finance*, 88, 113-128.
- Li, Y., Eddie, I., & Liu, J. (2014). Carbon emissions and the cost of capital: Australian evidence. *Review of Accounting and Finance*, 13(4), 400-420.
- Liu, P. H., Wang, Y., Xue, D. R., Linnenluecke, P. M., & Cai, D. C. W. (2021). Green Commitment and Stock Prick Crash Risk. *Finance Research Letters*.
- Monasterolo, I., & de Angelis, L. (2020). Blind to carbon risk? An analysis of stock market reaction to the Paris Agreement. *Ecological Economics*, 170.
- Murata, R., & Hamori, S. (2021). ESG disclosures and stock price crash risk. *Journal of Risk and Financial Management*, 14(2), 70.
- Oliveira, N. R. d. A. (2023). *Optimal portfolios based on stock characteristics: construction and performance for the US market* Universidade do Minho]. RepositóriUM. <https://hdl.handle.net/1822/84043>

- Perugini, L., Pellis, G., Grassi, G., Ciais, P., Dolman, H., House, J. I.,...Peylin, P. (2021). Emerging reporting and verification needs under the Paris Agreement: How can the research community effectively contribute? *Environmental Science & Policy*, 122, 116-126.
- Pham, H. N. A., Ramiah, V., Moosa, N., Huynh, T., & Pham, N. (2018). The financial effects of Trumpism. *Economic Modelling*, 74, 264-274.
- Phillips, R. L., & Ormsby, R. (2016). Industry classification schemes: An analysis and review. *Journal of Business & Finance Librarianship*, 21(1), 1-25.
- Qian, W., & Schaltegger, S. (2017). Revisiting carbon disclosure and performance: Legitimacy and management views. *The British Accounting Review*, 49(4), 365-379.
- Schmidt, P. S., von Arx, U., Schrimpf, A., Wagner, A. F., & Ziegler, A. (2019). Common risk factors in international stock markets. *Financial Markets and Portfolio Management*, 33, 213-241.
- Stavins, R. N. (2019). Carbon taxes vs. cap and trade: Theory and practice. *Cambridge, Mass.: Harvard Project on Climate Agreements*.
- Trinks, A., Ibikunle, G., Mulder, M., & Scholtens, B. (2022). Carbon intensity and the cost of equity capital. *The Energy Journal*, 43(2), 181-214.
- Venturini, A. (2022). Climate change, risk factors and stock returns: A review of the literature. *International Review of Financial Analysis*, 79.
- Yildiz, Y., & Karan, M. B. (2020). Environmental policies, national culture, and stock price crash risk: Evidence from renewable energy firms. *Business Strategy & the Environment (John Wiley & Sons, Inc)*, 29(6), 2374-2374-2391.
- Zaman, R., Atawnah, N., Haseeb, M., Nadeem, M., & Irfan, S. (2021). Does corporate eco-innovation affect stock price crash risk? *The British Accounting Review*, 53(5), 101031.

Appendices

A. Data Variables Description: Datastream Datatypes

This Appendix provides a comprehensive list of all Datastream datatypes, including those used in the filter implementations described in section 4, as well as the datatypes used in calculating some of the variables of control.

Datatype	Name	Definition	Type
TYPE	Type of Instrument	Mandatory classification assigned to all instruments.	Static
DSCD	Datastream Code	The unique six-digit quote identifier issued by Datastream.	Static
BDATE	Base Date	The timeseries history start date for a quote.	Static
ENAME	Expanded Security Name	The expanded form of the quote name.	Static
EXMNEM	Datastream Exchange Mnemonic	It follows the ISO standard and is represented by a 3-character alphabetic code. This code helps identify the source of default price datatypes for a particular equity.	Static
GEOGN	Geographical Classification of company	It specifies the home or listing country of a company security.	Static
ISIN	ISIN	The code that uniquely identifies a security and it was developed by the ISO as an international coding system.	Static
ISIND	Primary Indicator Flag	It shows either P or S. P means the main listing (in the home country), and S means a	Static

		secondary listing (in a foreign country)	
LOC	Code Local	It is an ID code using the official exchange code. It has up to 12 characters, starting with a letter-based country code.	Static
PCUR	Currency	The currency used to quote and display the security's price.	Static
TRAC	Security Type Code	It represents the category of share, and it is the Refinitiv devised classification for instruments.	Static
ESTAT	Equity Status	The trading status of the quote.	Static
TIME	Time – Latest Value	The date of the latest recorded close price.	Static
RI	Total Return Index	This describes the increase in the value of a 100-unit investment denominated in the local currency on the base date. It assumes that dividends are reinvested to acquire additional units of an equity or unit trust at the closing price on the ex-dividend date.	Time Series
MV	Market Value (Capital)	The share price multiplied by the number of outstanding ordinary shares.	Time Series

B. Data Variables Description: ESG Datatypes

This Appendix offers a detailed description of the ESG datatype used to calculate the carbon emissions measure discussed in Section 3.2.

Datatype	Name	Definition	Type
ENERDP024	CO2 Equivalent Emissions Direct, Scope 1	Direct CO2 and CO2 equivalents emission in tonnes. The following gases are relevant: carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorinated compound (PFCS), sulfur hexafluoride (SF6), nitrogen trifluoride (NF3).	Time Series

C. Data Variables Description: Worldscope Datatypes

This Appendix provides a comprehensive list of all Worldscope datatypes used in calculating the control variables.

Datatype	Name	Definition	Type
WC01001	Net Sales or Revenues	It represents gross sales and other operating revenue less discounts, returns and allowances.	Time Series
WC07021	Sic Code 1	SIC codes, classify economic activities across the United States. A company can have up to 8 SIC codes, with SIC code 1 indicating the segment generating the most revenue and SIC code 8 representing the least. When sales breakdown is unavailable, Worldscope assigns SIC codes at its discretion.	Static
WC05301	Common Shares Outstanding	It represents the number of shares outstanding at the company's year-end.	Time Series
WC08006	Trading Volume – Annual – US and Canada	Common Shares Traded-Annual * ((Market Price-High + Market Price-Low / 2) If High and Low Prices are not available, then Market Price - Year End is used.	Time Series
WC02999	Total Assets	Sum of total current assets, long term receivables, investment in unconsolidated subsidiaries, other investments, net property plant and equipment and other assets.	Time Series

WC03351	Total Liabilities	All short- and long-term obligations expected to be satisfied by the company.	Time Series
WC04101	Deferred Income Taxes & Investment Tax Credit	The increase or decrease in the deferred tax liability from one year to the next resulting from timing differences in recognition of revenues and expenses for tax and financial reporting purposes.	Time Series
WC03451	Preferred Stock	A claim prior to the common shareholders on the earnings of a company and on the assets in the event of liquidation.	Time Series
WC03251	Long-Term Debt	All interest-bearing financial obligations, excluding amounts due within one year. It is shown net of premium or discount.	Time Series
WC08326	Return On Assets	$\frac{(\text{Net Income} - \text{Bottom Line} + ((\text{Interest Expense on Debt} - \text{Interest Capitalized}) * (1 - \text{Tax Rate})))}{\text{Average of Last Year's and Current Year's Total Assets}}$	Time Series

D. Industry Classification of Sample Firm

This Appendix presents a breakdown of the 1,279 US-listed stocks by industry, classified according to the Fama-French 48 industry categories described in section 4. Accordingly, financials (SIC codes 6000-6999) and utilities (SIC codes 4900-4949) are excluded from this classification.

Industry Classification	Number of Firms	% of the Total
34 Business Services	204	15,95%
36 Electronic Equipment	89	6,96%
42 Retail	81	6,33%
30 Petroleum and Natural Gas	78	6,10%
40 Transportation	60	4,69%
21 Machinery	60	4,69%
14 Chemicals	57	4,46%
13 Pharmaceutical Products	48	3,75%
41 Wholesale	44	3,44%
23 Automobiles and Trucks	42	3,28%
43 Restaurants, Hotels, Motels	38	2,97%
12 Medical Equipment	37	2,89%
35 Computers	34	2,66%
32 Communication	34	2,66%
37 Measuring and Control Equipment	32	2,50%
2 Food Products	31	2,42%
17 Construction Materials	27	2,11%
9 Consumer Goods	25	1,95%
19 Steel Works Etc	25	1,95%
18 Construction	23	1,80%
38 Business Supplies	20	1,56%
22 Electrical Equipment	19	1,49%
10 Apparel	15	1,17%
28 Non-Metallic and Industrial Metal Mining	14	1,09%
33 Personal Services	14	1,09%
4 Beer & Liquor	13	1,02%
24 Aircraft	13	1,02%
15 Rubber and Plastic Products	12	0,94%
6 Recreation	10	0,78%
8 Printing and Publishing	10	0,78%

39 Shipping Containers	10	0,78%
11 Healthcare	9	0,70%
7 Entertainment	7	0,55%
48 Other	7	0,55%
3 Candy & Soda	6	0,47%
29 Coal	6	0,47%
1 Agriculture	4	0,31%
5 Tobacco Products	4	0,31%
20 Fabricated Products	4	0,31%
27 Precious Metals	4	0,31%
16 Textiles	4	0,31%
25 Shipbuilding, Railroad Equipment	3	0,23%
26 Defense	2	0,16%
Total	1279	100,00%