Essays on Measures of Risk in Finance

Thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

ICMA CENTRE

HENLEY BUSINESS SCHOOL

THE UNIVERSITY OF READING

Jingqi Pan

September 2023

Imagination is more important than knowledge.

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Jingqi Pan

Acknowledgements

I am immensely thankful to my advisors, Dr Emese Lazar and Dr Shixuan Wang, for their guidance during my PhD journey. Emese has nurtured my academic growth, teaching me to think critically and excel in research writing. Shixuan has constantly encouraged me to overcome challenges and aspire for excellence. Their mentorship has been pivotal in my development as a researcher, and I could not have reached this milestone without them. I would also like to thank Dr Haoshen Hu, for the enlightening discussions.

I sincerely thank the ICMA Centre for its generous financial support. I am very grateful to my family, friends, and colleagues for being very kind and supportive.

Abstract

This thesis explores financial risk measurement, specifically addressing market risk, climate transition risk, and credit risk, which are organized into three main chapters.

The first contribution is that it proposes to estimate Value-at-Risk (VaR) and Expected Shortfall (ES) at extreme levels using information from a common level. We employ time series cross-validation to optimize the forecasting performance. Our simulation study reveals that the proposed novel models outperform the original Generalized Autoregressive Score model of Patton et al. (2019) according to various backtests. Empirical evidence based on the return data of four oil futures also shows the superior performance of the proposed models. Notably, our models' performance is most prominent during the COVID-19 period.

The second contribution is that it provides a framework to measure the effects of climate transition risk factors, proxied by the environmental pillar of ESG scores, on corporate downside risk. Analyzing the stock returns and climate risk factors relationship, a notable negative correlation in lower quantiles is revealed. A new risk measure for climate transition risk factors is also proposed, with empirical findings indicating sector-based variations in sensitivity to these risks. Specifically, the Health Care sector is the least efficient in reducing climate

risk, while the Energy sector benefits the most from improvements in the firms' environmental scores.

The third contribution is a study examining how corporate environmental performance influences credit ratings, with a trans-Atlantic study encompassing firms from the United States (US) and the European Union (EU). We find that corporate environmental performance positively affects the firms' credit ratings. Interestingly, our findings reveal a linear relationship in the US and a nonlinear one in the EU. These findings highlight the implications of environmental performance. They provide vital insights for firms aiming to improve their credit rating via sustainability initiatives, while considering regional disparities.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation for the Thesis

In the recent two decades, there is an increasingly complex financial landscape, and the importance of risk management is more profound than ever. Understanding, measuring, and managing risks are vital in maintaining financial stability, especially in an era characterized by technological advancements, and growing concerns about environmental sustainability. Risk measurement tools provide insight into potential losses and offer strategies to forecast and manage such risks.

The notion of risk originates from the critical intersection of uncertain future outcomes and the potential for undesirable results. When navigated successfully, risk management can protect stakeholders' interests, promote financial stability, and encourage sustainable growth. The global financial landscape is in a state of constant evolution, shaped by numerous and varied risks.

In recent decades, we have witnessed catastrophic financial events, such as the global financial crisis of 2008 and the COVID-19 market crash. These events highlight the necessity of robust methodologies for measuring risk. Concurrently, as awareness of global warming intensifies over time, it is increasingly evident that corporations and investors can no longer overlook the importance of climate risks in their decision-making processes. Moreover, as sustainability and corporate social responsibility gain prominence in strategic planning, the connection between environmental performance and credit risk has garnered considerable attention.

Risk has been traditionally categorized according to multiple aspects, reflecting various uncertainty sources that could negatively affect a company or the entire financial market. Among all categories of risks, the thesis focuses on market risk, climate risk, and credit risk. Each type of risk carries its unique characteristics and tools, often necessitating specialized strategies for effective management.

Market risk refers to the potential loss in the value of investments due to fluctuations in market factors like interest rates, exchange rates, and equity prices. This risk can negatively impact individual investments, portfolios, and the broader financial system. The statistical risk measures, Value-at-Risk (VaR) and Expected Shortfall (ES) are widely accepted for market risk measurement and management. VaR estimates the maximum loss that could occur with a given confidence level over a given period. As required by the Basel Committee on Banking Supervision (2013), market risk is typically measured by ES which estimates the average loss when VaR is exceeded. Given significant market downturns, such as the global financial crisis and the COVID-19 market crash, the ability to accurately measure and manage market risk has become exceedingly important. This demand has led to the development of innovative models, like the proposed extensions to the one-factor GAS model, aimed at estimating risk at extreme levels.

Another evolving and pressing category of risk is climate risk, stemming from the changing environmental conditions due to global warming. In financial terms, climate risk refers to potential losses arising from shifts in climate patterns. This includes physical risks that are a direct result of environmental changes such as extreme weather events, as well as transition risks that are tied to the economic adjustments society needs to make towards transitioning to a low-carbon economy. In recent years, the financial industry has seen an increasing need to quantify and manage climate risk, driven by both a heightened regulatory focus and rising investor interest in sustainable investments that consider climate impacts. This growing demand has turned climate risk into a crucial factor that can affect companies' downside risk, which is the potential for companies to experience financial losses. In simpler terms, as our environment changes, these changes can directly influence how much a company might stand to lose during difficult times. As evidenced in our investigation into climate VaR and climate ES, it is apparent that climate risk is not exclusive to energy companies alone. Rather, it affects all sectors and can significantly impact the total risk of equity.

Credit risk refers to the potential for financial loss if a borrower fails to repay debt or defaults on contractual obligations. Measurement and management of credit risk is crucial for financial stability, and numerous approaches exist to quantify this risk. An important method is the use of credit ratings from independent rating agencies such as Standard & Poor's (S&P), Moody's, and Fitch Ratings. These agencies evaluate and assign ratings that reflect the creditworthiness of debt issuers such as corporations and governments. Ratings typically range from "AAA" or "Aaa" for low-risk borrowers to "D" for those already in default.[1] Another widely employed measure of credit risk is the Probability of Default (PD), which quantifies the probability that firms financed by debt capital may fail to fulfill their debt obligations through the repayment of principal and interest at the specified maturity periods (Vassalou and Xing, 2004). PD is typically derived using statistical models that factor in various risk determinants, such as a borrower's financial health, market conditions, and macroeconomic information. A complementary measure, often used in tandem with PD, is the Distance-to-Default (DtD) a metric proposed by Merton (1974). DtD is a market-based measure of a firm's credit risk that quantifies how far a firm is from default. Unlike credit ratings and PD, which are often based on historical data and subjective assessments, DtD is derived from real-time market data, which is a more forward-looking indicator of credit risk. It should also be noted that these traditional credit risk measures are increasingly being integrated with sustainability factors, as demonstrated by our study on the relationship between environmental performance and credit ratings. Interestingly, this relationship differs across geographies, indicating that regional differences also play a role in credit risk assessment.

In summary, the management of these risks (market risk, climate risk, and credit risk) is not just a matter of prudent business practice; it also contributes to broader economic stability and sustainability. Understanding the importance of measuring these diverse types of risk can lead to the development of more effective risk management strategies.

1.2 Overview of the Thesis

Chapter 2 presents a new framework for jointly modeling and forecasting dynamic VaR and ES at extreme percentiles. This is achieved by simultaneously estimating VaR and ES at two different levels of significance, namely at an extreme level and at an auxiliary level in the semiparametric models. This innovative approach is developed from the semiparametric models introduced by Patton et al. (2019) for forecasting VaR and ES jointly based on the Generalized Autoregressive Score (GAS) framework of Creal et al. (2013). The GAS framework is an observation-driven model that updates the time-varying parameters based on a scaled score. To evaluate the improvement of integrating information from the auxiliary level into the GAS framework, we employ Time Series Cross-Validation (TSCV) to select the optimal auxiliary level within the range of 2.5% - 20%. Therefore, Chapter 2 of this thesis sheds light on extensions of the GAS model and the improvement of forecasting performance at extreme levels by utilizing models that incorporate information from an auxiliary level.

Our simulation study contrasts our Augmented GAS (A-GAS) models with Patton et al. (2019)'s GAS models using Monte Carlo simulations. We employ the GJR-GARCH (1,1) model of Glosten et al. (1993) with skewed t distribution as the data generation process (DGP). To compare the performance of the risk models, we employ backtests for VaR or ES forecasts individually. The backtests we consider include the unconditional coverage test introduced by Kupiec (1995) and the conditional coverage test proposed by Christoffersen (1998) for VaR and the bootstrap test of McNeil and Frey (2000) for ES. Regarding the joint (VaR, ES) backtests, we compare the average loss values generated by the FZ0 loss function proposed by Fissler and Ziegel (2016). The simulation results show that our A-GAS models outperform the original GAS model, where the A-GAS models with TSCV perform the best among different auxiliary levels.

In the empirical study, our A-GAS models are applied to four commodities futures and compared with a range of parametric, nonparametric, and semiparametric models, including historical simulations, GARCH, and the original GAS models. We base the comparison on the same backtesting methods for VaR and ES as in the simulation study. Moreover, we employ the Diebold-Mariano test of Diebold and Mariano (2002) to compare the relative performance between models.

Beyond extreme market crash losses, financial regulators and investors are growing more concerned about the effect of climate change on investments, aiming to gauge the associated risks. Whilst energy companies have attracted most of the attention due to the contribution of the Energy sector to climate change, climate risk actually affects companies in every sector. To echo this increasing concern, we explore how climate-related factors impact the firms' downside risk, as measured by VaR and ES in the following chapter.

Chapter 3 aims to quantify the impact of climate change on the market risk of equities across various sectors. We introduce novel measures, namely climate Value-at-Risk (climate VaR) and climate Expected Shortfall (climate ES), which are used to capture the risk attributed to transition climate risk factors proxied by environmental scores. The environmental scores represent the firms' commitment to reducing their environmental impact and are obtained from the Refinitiv ASSET4 database.

Our first finding reveals a significantly negative relationship between stock returns and transition climate risk factors in the lower quantiles of stock returns. We also find significant heterogeneity in the sensitivity of firm-level downside risk to environmental scores, with some sectors benefiting from improvements in environmental scores and others experiencing an increase in loss. For instance, the Energy sector appears to benefit the most from improvements in environmental scores, while the Health Care sector sees an increase in firms' total downside risk as their environmental scores improve.

We also introduce a new concept, that of the "climate risk ratio", which indicates the extent to which the environmental scores affect the total downside risk of the firms. The sectors like Basic Materials, Consumer Staples, Energy, Financials, Technology, and Utilities benefit from efforts to increase the companies' environmental scores, while sectors like Consumer Discretionary, Health Care, Industrials, Real Estate, and Telecommunications see an increase in their total downside risk with an improvement in their environmental scores.

Different stakeholders may prioritize various risk types. Shareholders and potential investors might be more interested in the downside risk attributed to environmental factors, while banks and other lenders might value more the impact of environmental factors on the credit risk of the company. Moreover, downside risk and credit risk are often correlated. Companies with higher market risk might appear less stable, possibly facing increased credit risk as lenders adjust interest rates for this added volatility.

Therefore, Chapter 4 investigates the impact of firms' environmental performance on their credit ratings in the US and EU. We first hypothesize that firms with better environmental performance, as measured by environmental scores from the Thomson Reuters ASSET4 ESG database, would have higher credit ratings. The credit ratings measures are the numerical conversion of the long-term foreign currency issuer credit ratings of S&P, Moody's, and Fitch. Our findings show a positive and significant relationship between firms' environmental performance and their credit ratings in both the US and EU.

The influence of environmental performance on credit ratings is more pronounced in the US than in the EU. Our study shows that the gap in credit ratings between environmentally efficient and inefficient firms is wider in the US. We delved deeper to understand the reasons behind this regional variation. By examining the distribution of environmental scores in both regions, we observed that a majority of US firms lag in environmental performance, while EU firms generally perform well. Notably, the EU presents a more nonlinear relationship between credit ratings and environmental performance. To ensure the reliability of our study, several endogeneity and robustness tests are implemented, reinforcing the positive relationship between environmental performance and credit ratings.

1.3 Original Contributions

This thesis, consisting of the following three main chapters, contributes to measuring different financial risks including market risk, climate risk, and credit risk:

(1) The first set of original contributions in terms of market risk is:

- we propose novel semiparametric models to jointly forecast VaR and ES at extreme significance levels by incorporating information from a common level into the GAS framework;

• we illustrate the application of TSCV to optimize the forecast performance of these models;

• we present evidence of the superior performance of our models against benchmarks based on Monte Carlo simulations using several backtesting methods;

• we provide empirical evidence showing the dominance of our augmented semi-parametric models over other benchmarks in oil futures data;

• we perform model comparisons during the COVID-19 pandemic.

(2) The second set of original contributions in terms of climate risk is:

• we highlight the relationship between stock returns and climate transition risk factors across different quantiles;

• we introduce innovative climate risk measures to capture VaR and ES associated with transition climate risk factors;

• we document sector-specific variations in the sensitivity of VaR and ES to environmental scores;

• we demonstrate the statistical and economic significance of environmental scores for VaR;

• we apply our methodology across different risk models, consistently reaffirming our findings.

(3) Our third set of original contributions in terms of credit risk is as follows:

• we examine the relationship between environmental performance and credit ratings and compare the differences between the US and EU using latest data from three credit rating agencies;

• we implement the numerical conversion of credit ratings into a 58-point system that takes into account the outlook and watch signals;

• we observe a nonlinear relationship between environmental performance and credit ratings in the EU;

• we pioneer the idea that regional environmental performance nuances might influence the relationship between corporate environmental performance and credit risks;

• we discover that the EU's nonlinear relationship is specific to environmental performance and not a general characteristic of ESG scores.

1.4 Outline of the Thesis

The structure of the remaining chapters is as follows: Chapter 2 studies the A-GAS models, integrating information from auxiliary levels for VaR and ES forecasting; Chapter 3 proposes measures to capture the impacts from climate transition risk factors to firm-level VaR and ES; Chapter 4 investigates the relationship between environmental performance and credit ratings in the US and EU. Chapter 5 summarizes the main discoveries and points toward avenues for subsequent research based on the findings of this thesis.

For a seamless reading experience, each chapter can be considered independent reading. We (re)introduce variables and abbreviations in each chapter. Whenever possible, we endeavour to follow consistent notations throughout this thesis.

Notes

[1]For Moody's credit ratings, "C" represents default.

Chapter 2

On the estimation of Value-at-Risk and Expected Shortfall at extreme levels

2.1 Introduction

Many institutional decisions in financial risk management, such as those related to capital requirements, rely on good forecasts of conditional distributions of asset returns, with an emphasis on the left tails of these distributions. What keeps risk managers awake at night are not typical price fluctuations but unexpected downfalls of unusual magnitudes. The concern is that these may trigger systemic spirals that can cause big losses. Financial regulators are concerned with protecting the financial system against catastrophic events that could be a source of systemic risk. It is of interest to correctly measure risk at very small levels of significance, but the small number of observations in the extreme tail of the returns' distribution constitutes a problem that such extreme returns occur very rarely. For daily returns, by definition, events that breach the 1% quantile occur about twice a year. Returns in the more extreme quantiles occur even more rarely, and our focus is the risk assessment of such events.

In this paper, we propose a framework to measure risk at extreme percentiles that extends two models of Patton et al. (2019) by simultaneously estimating risk at two different levels of significance (an extreme level and an auxiliary level), by assuming a joint process that drives both sets of risk measures. The optimal choice of auxiliary level is a more common level (in the range of 2.5% - 20%) which can be selected via time series cross-validation. We illustrate via simulations and commodities data that by simultaneously considering an auxiliary level of significance, the risk estimates at the extreme levels of significance outperform the alternatives in terms of loss values, and often in terms of backtest performance as well.[1]

Value-at-Risk (VaR) is one of the most popular tail risk measures that is employed to assess and manage financial risk. VaR is an estimate of the quantile of the distribution of profit and losses and it can be measured at different significance levels. Due to its conceptual simplicity, VaR has become a popular risk measure of market risk. However, VaR ignores the shape and structure of the tail of the returns' distribution and is not a coherent risk measure (i.e. it is not "sub-additive") (Artzner et al., 1999). Expected Shortfall (ES) is a risk measure that has recently increased in popularity due to its favourable properties. It measures the expected value of the observations provided that they exceed VaR and is a coherent risk measure (Roccioletti, 2015). A transition from VaR at 1% level to ES at 2.5% level has been proposed by the Basel Committee on Banking Supervision (2013). However, the measurement of ES is inherently dependent on the value of the VaR estimate. As such, ES is not elicitable by itself, and only the (VaR, ES) tuple is elicitable (Ziegel, 2016).

It is well known that the volatility displayed by commodity market returns has often been high (Hung et al., 2008) - as also shown by the recent events related to COVID-19 and the ongoing international conflict between Russia and Ukraine. It has been documented that commodity asset returns are generally characterized by higher volatility than stock returns (Del Brio et al., 2020). Thus, it is vital to have special risk management tools for the commodity market, which are needed by market participants and policymakers. Specifically, market participants need to measure market risk at extreme levels in order to manage their portfolios. As for policymakers, they need to be aware of the risks faced by the economy, because extreme commodity price changes can have a big impact on the economy as a whole, as indicated by Sadorsky (1999) and others.

The literature on VaR and ES estimation is very rich. To measure risk at multiple significance levels, White et al. (2015) propose a vector autoregressive (VAR) framework to quantile models which extend the CAViaR model of Engle and Manganelli (2004) to multiple confidence levels. Following the results of Fissler and Ziegel (2016) that ES and VaR are jointly elicitable, Patton et al. (2019) present several novel models. Specifically, they propose four dynamic semiparametric models for VaR and ES, based on the generalized autoregressive score (GAS) framework (see Creal et al., 2013; Harvey, 2013). However, VaR and ES at the popular significance levels (e.g. 1%, 2.5%, and 5%) provide insufficient information about rare but drastic events such as the COVID-19 crisis. Also, copula models can be used to improve VaR predictions, such as in Li et al. (2022). Many papers ignore the possibility of multiple regimes in the risk models; one way to address this problem is by using Markov Switching models, as in Maciel (2021).

Researchers have devoted effort to estimate VaR and ES at extreme significance levels. There is no well-defined definition of extreme level for risk, but in the literature it is typically defined as at significance levels below 1%. Chavez-Demoulin et al. (2014) propose a nonparametric extension of the Peaks-Over-Threshold method from Extreme Value Theory (EVT) to estimate VaR and ES at 1% significance level. Hoga (2017) proposes tests to detect changes in extreme VaR at significance levels below 1% based on the Weissman estimator motivated by EVT. Danielsson and De Vries (1998) propose a semi-parametric method to assess the probability of extreme events for data with heavy tails and apply it for VaR at extreme significance levels such as 0.5%, 0.1%, and 0.005%. In the study of Kourouma et al. (2010), VaR and ES are estimated based on the EVT model using the Peaks-Over-Threshold method, and it is shown that this type of EVT model performs better during the 2008 financial crisis than the unconditional VaR models. Based on a GARCH-type volatility model with covariates, Hoga (2021) derives asymptotically valid forecast intervals for VaR and ES, which are proved to be adequate for extreme risk levels. The above papers all focus on VaR and ES estimations at extreme levels of significance, but whilst they are based on EVT, our models forecast risk measures based on the GAS framework. There are several papers that improve on risk forecasts via forecast combinations, such as Taylor (2020) and Storti and Wang (2022), and the latter proposes forecast combinations of VaR models for various quantiles in order to compute ES. Our approach is, however, to use the information from a specific generic quantile to improve VaR and ES forecasts at an extreme level of significance.

This paper makes three main contributions. First, from a methodology perspective, we propose an extension of two models (the one-factor GAS model and the hybrid GAS/GARCH model) of Patton et al. (2019) to be used for risk estimation at extreme levels of significance, by simultaneously estimating VaR and ES at two different levels of significance, namely at an extreme level and at an auxiliary level. Without relying on such an auxiliary level, the extreme risk measure will depend on a small number of observations in the extreme tail of the empirical distribution of the returns. Therefore, incorporating information on a more generic tail can help to improve the forecast of VaR and ES at extreme levels. We obtain parameter estimates that are more robust than the parameters of standard GAS models, as highlighted by our simulations. Second, from a practical perspective, we demonstrate how to employ time series cross-validation (TSCV) to select the optimal auxiliary level from a set of candidates in order to improve the forecast performance of the proposed models. The TSCV is a data-driven method that helps with the selection of the auxiliary level without relying on arbitrary judgment. Third, from an empirical perspective, we provide compelling evidence that our models outperform the alternatives in terms of the evaluation of VaR and ES forecasts in a forecasting exercise. Our empirical analysis is based on four oil futures and we find that the recent COVID-19 crisis period well illustrates the strengths of our models in terms of forecasting risk at extreme levels.

The rest of the paper is organized as follows. Section 2.2 discusses VaR and ES models including the four GAS models proposed by Patton et al. (2019) and introduces the proposed GAS models that simultaneously estimate VaR and ES at two levels of significance. The simulation results regarding model performance are presented in Section 2.3. Section 2.4 presents the data used in our empirical study, the in-sample estimation results, and out-of-sample (OOS) forecast results. Section 2.5 presents robustness results based on a rolling window estimation. Section 2.6 concludes. An online Supplemental Appendix provides additional results.

2.2 The augmented GAS model

2.2.1 Modelling VaR and ES

VaR provides banks and financial institutions with an estimate of the minimum loss level that occurs in the worst outcomes at a given significance level $\alpha \in (0, 1)$. Let $F_Y(\cdot|\Omega_{t-1})$ denote the cumulative distribution function of asset return Y_t over a time horizon (such as one day or one week) conditional on the information set Ω_{t-1}. Following Ziegel (2016), Nolde and Ziegel (2017), and Chen (2018), the VaR at level α at time t can be defined as:

$$VaR_t^\alpha = \inf\{Y_t|F_Y(Y_t|\Omega_{t-1}) \geq \alpha\}, \tag{2.2.1}$$

where VaR_t^α denotes the α-quantile of the underlying return distribution at time t. As such, VaR at level α can be written directly in terms of the inverse cumulative distribution function (Duffie and Pan, 1997):

$$VaR_t^{\alpha} = F_Y^{-1}(\alpha|\Omega_{t-1}). \tag{2.2.2}$$

ES measures the expectation of returns conditional on their value being less than VaR. ES is a coherent risk measure (Roccioletti, 2015) due its superior properties, and it has become increasingly popular in the risk management of banks and financial institutions. Recently, the Basel Committee on Banking Supervision (2013) proposed a transition from VaR at 1% level to ES at 2.5% level motivated by the global financial crisis in 2008. ES at level α at time t can be formally defined as (see Acerbi and Tasche, 2002):

$$ES_t^{\alpha} = \mathbb{E}[Y_t|Y_t \leq VaR_t^{\alpha}, \Omega_{t-1}]. \tag{2.2.3}$$

2.2.2 Generalized Autoregressive Score (GAS) framework

The application of the GAS framework for VaR and ES forecasting has been introduced by Patton et al. (2019). They propose the two-factor GAS model, the one-factor GAS model, the GARCH-FZ model, and the hybrid GAS/GARCH model to estimate VaR and ES jointly by minimizing the expectation of the VaR and ES joint loss function.[2] One of the most popular loss functions is the FZ0 loss function proposed by Fissler and Ziegel (2016), which has been further popularized by Patton et al. (2019). The FZ0 loss function is defined as:

$$L_{FZ0}(Y, v, e; \alpha) = -\frac{1}{\alpha e}\mathbf{1}\{Y \leq v\}(v - Y) + \frac{v}{e} + log(-e) - 1, \tag{2.2.4}$$

where Y is the return on the underlying asset, and v and e denote VaR and ES, respectively. $\mathbf{1}\{Y \leq v\}$ is an indicator function which returns 1 when $Y \leq v$ (i.e., the VaR is exceeded). Loss differences generated from the FZ0 loss function are homogeneous of degree zero. When $Y > v$, the returns do not affect the value of the loss. However, the loss value heavily relies on the returns when $Y \leq v$, with the parameter estimates being influenced by these extreme returns through the score. The parameters of the GAS models of Patton et al. (2019) are estimated by minimizing the loss function in Eq.(2.2.4). In the following, we briefly summarize their four model specifications.

The two-factor GAS model for ES and VaR

In the two-factor GAS model, the forecasts of VaR and ES are determined by the current value of VaR and ES and the forcing variable which is a function of the first order derivative and the Hessian of L_{FZ0}. The specification of the two-factor GAS model is shown below:

$$\begin{bmatrix} v_{t+1} \\ e_{t+1} \end{bmatrix} = \mathbf{W} + \mathbf{B} \begin{bmatrix} v_t \\ e_t \end{bmatrix} + \mathbf{A}\mathbf{H}_t^{-1}\nabla_t, \tag{2.2.5}$$

where $\mathbf{W}$ is a (2×1) vector and $\mathbf{B}$ and $\mathbf{A}$ are (2×1) matrices. The scoring function is given by:

$$\nabla_t \equiv \begin{bmatrix} \partial L_{FZ0}(Y_t, v_t, e_t; \alpha)/\partial v_t \\ \partial L_{FZ0}(Y_t, v_t, e_t; \alpha)/\partial e_t \end{bmatrix}, \tag{2.2.6}$$

and the scaling matrix $\mathbf{H}_t$ is the Hessian matrix:

$$\mathbf{H}_t = \begin{bmatrix} \frac{\partial^2 \mathbb{E}_{t-1}[L_{FZ0}(Y_t,_t,e_t;\alpha)]}{\partial v_t^2} & \frac{\partial^2 \mathbb{E}_{t-1}[L_{FZ0}(Y_t,v_t,e_t;\alpha)]}{\partial v_t \partial e_t} \\ . & \frac{\partial^2 \mathbb{E}_{t-1}[L_{FZ0}(Y_t,v_t,e_t;\alpha)]}{\partial e_t^2} \end{bmatrix}. \tag{2.2.7}$$

The one-factor GAS model for ES and VaR

The two-factor model allows ES and VaR to be updated as two separate, but correlated, processes. However, in the one-factor GAS model, VaR and ES are based on a time-varying risk measure κ_t (similar to the conditional variance process in the GARCH model). The one-factor GAS model is written as:

$$\begin{aligned} v_t &= a \ \exp\{\kappa_t\}, \\ e_t &= b \ \exp\{\kappa_t\}, \ b < a < 0, \\ \kappa_t &= \omega + \beta\kappa_{t-1} + \gamma H_{t-1}^{-1} s_{t-1}, \end{aligned} \tag{2.2.8}$$

where the restriction $b < a < 0$ follows Patton et al. (2019) and s_t is given by:

$$s_t \equiv \frac{\partial L_{FZ0}(Y_t, a\exp\{\kappa_t\}, b\exp\{\kappa_t\}; \alpha)}{\partial \kappa_t} = -\frac{1}{e_t}\left(\frac{1}{\alpha}\mathbf{1}\{Y_t \leq v_{t-1}\}Y_t - e_t\right), \tag{2.2.9}$$

and for simplicity, Patton et al. (2019) set the Hessian factor H_t as one. Thus the one-factor GAS model for ES and VaR can be written as:

$$\kappa_t = \omega + \beta\kappa_{t-1} + \gamma\frac{1}{b\exp\{\kappa_{t-1}\}}\left(\frac{1}{\alpha}\mathbf{1}\{Y_{t-1} \leq a\exp\{\kappa_{t-1}\}\}Y_{t-1} - b\exp\{\kappa_{t-1}\}\right). \tag{2.2.10}$$

The GARCH-FZ model for ES and VaR

Forecasting VaR and ES via a GARCH type model is one of the most prevailing ways to estimate risk measures, due to its parsimony. The GARCH-FZ model employs the framework of a GARCH model to generate VaR and ES, but the parameters of this model are estimated by minimizing the expectation of the loss function FZ0, instead of using (Q)MLE. The model is:

$$\begin{aligned} Y_t &= \mu_t + \sigma_t \, \eta_t, \quad \eta_t \sim iid \; F_\eta(0,1), \\ \sigma_t^2 &= \omega + \beta\sigma_{t-1}^2 + \gamma Y_{t-1}^2, \end{aligned} \tag{2.2.11}$$

where σ_t^2 is the conditional variance which follows a GARCH(1, 1) process. In terms of VaR and ES, the dynamic structure is analogous to the one-factor GAS model shown above:

$$\begin{aligned} v_t &= a \; \sigma_t, \quad \text{where } a = F_\eta^{-1}(\alpha), \\ e_t &= b \; \sigma_t, \quad \text{where } b = \mathbb{E}\left[\eta_t | \eta_t \leq a\right], \end{aligned} \tag{2.2.12}$$

where η_t is the standardized residual.

The hybrid GAS/GARCH model for ES and VaR

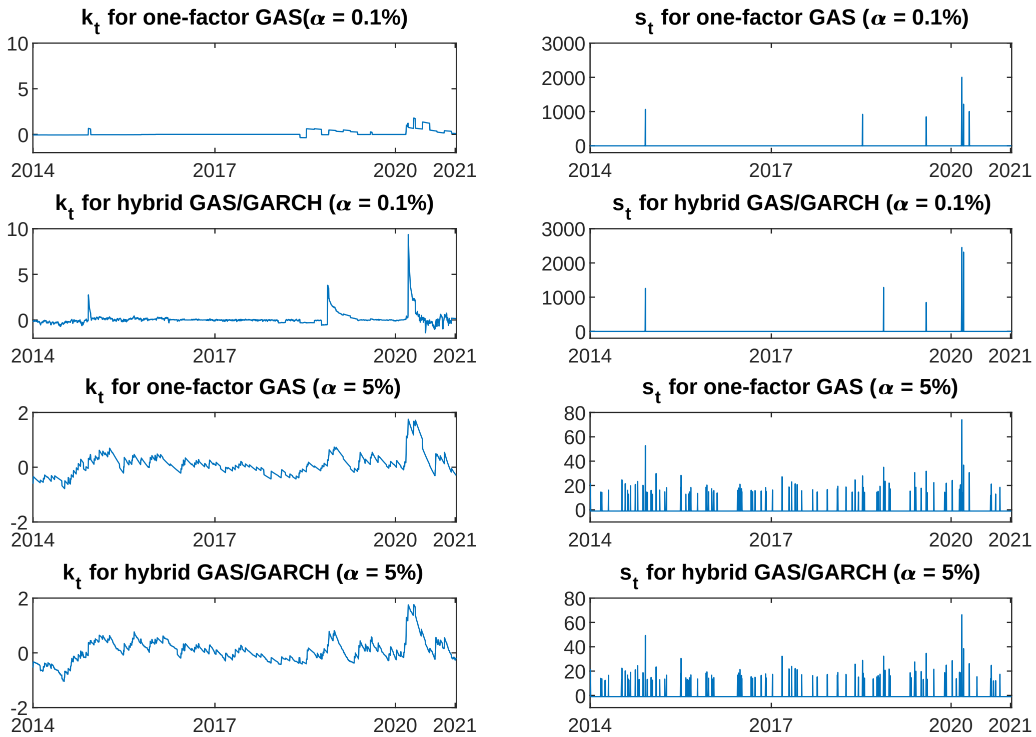
In the hybrid GAS/GARCH model, the process κ_t in the one-factor GAS model and the volatility σ_t in the GARCH model both contribute to the dynamics of VaR and ES. Thus, as a combination of both models, the hybrid GAS/GARCH model is specified as:

$$
\begin{aligned}
& Y_t = \exp\{\kappa_t\}\eta_t,\ \eta_t \sim iid\ F_\eta(0,1), \\
& v_t = a\ \exp\{\kappa_t\}, \\
& e_t = b\ \exp\{\kappa_t\},\ b < a < 0, \\
& \kappa_t = \omega + \beta\kappa_{t-1} + \gamma\frac{1}{e_{t-1}}\left(\frac{1}{\alpha}\mathbf{1}\{Y_{t-1} \leq, v_{t-1}\}Y_{t-1} - e_{t-1}\right) + \delta\log|Y_{t-1}|,
\end{aligned}
\tag{2.2.13}
$$

where κ_t is the log-volatility which is affected by Y_{t-1} in terms of the logarithm of absolute return rather than the square of return.

We now turn our attention to modelling risk at extreme levels within a GAS framework. According to Eq.(2.2.10) and Eq.(2.2.13), κ_t depends on s_{t-1} (the first order derivative of the FZ0 loss function, driven mostly by the indicator function in Eq.(2.2.4)) and κ_{t-1}. Figure 2.2.1 presents the κ_t and s_t processes for the one-factor GAS and hybrid GAS/GARCH models for two different significance levels, estimated from WTI crude futures prices. In general, κ_t remains mostly unchanged, and so the VaR and ES at time t are largely unaffected by the small changes in κ_t. For a less extreme significance level ($\alpha = 5\%$, for example), κ_t and s_t are more dynamic, being influenced by the returns in the tail of the distribution (see the last four figures in Figure 2.2.1). VaR and ES at a higher significance level can use past information more efficiently. Therefore, in order to improve on the estimation of GAS models of Patton et al. (2019) for extreme levels of significance, we propose the augmented GAS models, which are introduced in the following section.

Figure 2.2.1: κ_t and s_t processes of GAS models at 1% and 5% level



Note: This figure presents the κ_t and s_t processes of the GAS one-factor (left) and Hybrid (right) models for $\alpha = 0.1\%$ and $\alpha = 5\%$, estimated for the WTI futures prices from Jan 2014 to Jan 2021, with the model parameters re-estimated every 30 trading days using a rolling window of 1805 observations (7 years).

2.2.3 The augmented GAS models for ES and VaR

In this section, we propose to enhance two dynamic semi-parametric models, which are the one-factor GAS model and the hybrid GAS/GARCH model of Patton et al. (2019), to improve the forecasts of risk measures at an extreme significance level. We achieve this by simultaneously modelling VaR and ES at two different levels, an extreme level α_1 (such as 0.1%) and an auxiliary level α_2 (a more common level in the range of 2.5% - 20%).[3] In this setup, the same unique hidden process drives the risk estimates of VaR and ES for both levels. As such, we introduce two augmented GAS models, namely the augmented GAS one-factor model (we label it as A-GAS-1F) and the augmented hybrid model (we label it as A-Hybrid). These two models are jointly labeled as A-GAS models. We denote the VaR and ES at the extreme level of interest α_1 as $v_{1,t}$ and $e_{1,t}$, and at the auxiliary level α_2 as $v_{2,t}$ and $e_{2,t}$. Also, we investigate the backtesting performance of the $v_{1,t}$ and $e_{1,t}$ forecasts because these are at the level of interest. In the following, we elaborate on the details of the proposed models.

The augmented GAS one-factor model for ES and VaR

Under the GAS framework, the VaR and ES processes linearly depend on κ_t. Similarly, in the A-GAS-1F model, $v_{1,t}$, $e_{1,t}$, $v_{2,t}$ and $e_{2,t}$ are all driven by κ_t, which on the other hand depends on its lagged values (κ_{t-1}) and the score at the auxiliary level α_2. The model can be defined as:

$$
\begin{aligned}
v_{1,t} &= a_1 \exp\{\kappa_t\},\ e_{1,t} = b_1 \exp\{\kappa_t\},\\
v_{2,t} &= a_2 \exp\{\kappa_t\},\ e_{2,t} = b_2 \exp\{\kappa_t\},\\
\kappa_t &= \omega + \beta\kappa_{t-1} + \gamma s_{t-1,\alpha_2},\\
s_{t,\alpha_2} &\equiv \frac{\partial L_{FZ0}(Y_t, a_2 \exp\{\kappa_t\}, b_2 \exp\{\kappa_t\}; \alpha_2)}{\partial \kappa_t} = -\frac{1}{e_{2,t}}\left(\frac{1}{\alpha_2}\mathbf{1}\{Y_t \leq v_{2,t}\}Y_t - e_{2,t}\right),
\end{aligned}
\tag{2.2.14}
$$

where $v_{1,t}$ and $e_{1,t}$ are the VaR and ES at the extreme level α_1 and $v_{2,t}$ and $e_{2,t}$ are the VaR and ES at the auxiliary level α_2. The score s_{t,α_2} only depends on α_2, being the first order derivative of the FZ0 loss function for the auxiliary level α_2.

The augmented hybrid GAS/GARCH model for ES and VaR

Extending the hybrid GAS/GARCH model of Patton et al. (2019), we propose the augmented hybrid GAS/GARCH model (labeled A-Hybrid) which uses an auxiliary level of risk α_2, given by:

$$
\begin{aligned}
Y_t &= \exp\{\kappa_t\}\eta_t, \quad \eta_t \sim iid\ F_\eta(0,1),\\
v_{1,t} &= a_1 \exp\{\kappa_t\},\ e_{1,t} = b_1 \exp\{\kappa_t\},\\
v_{2,t} &= a_2 \exp\{\kappa_t\},\ e_{2,t} = b_2 \exp\{\kappa_t\},\\
\kappa_t &= \omega + \beta\kappa_{t-1} + \gamma\left(\frac{1}{e_{2,t-1}}\left(\frac{1}{\alpha_2}\mathbf{1}\{Y_{t-1} \leq v_{2,t-1}\}Y_{t-1} - e_{2,t-1}\right)\right) + \delta \log|Y_{t-1}|,
\end{aligned}
\tag{2.2.15}
$$

where the log-volatility κ_t is the same as in the hybrid GAS/GARCH model but is based on α_2.[4]

Parameter estimation

In the augmented models, the forecasts of risk measures at the extreme level α_1 consider the losses at the auxiliary risk level α_2. Thus, the VaR and ES at α_1 are obtained by minimizing the joint loss function which is the sum of both FZ0 loss functions, at both α_1 and α_2 levels.[5] Let $\mathfrak{L}_{FZ0}$ be the sum of the FZ0 loss functions, defined as:

$$\mathfrak{L}_{FZ0}(Y, v_1, v_2, e_1, e_2; \alpha_1, \alpha_2) = \underbrace{L_{FZ0}(Y, v_1, e_1; \alpha_1)}_{L_{FZ0}(\alpha_1)} + \underbrace{L_{FZ0}(Y, v_2, e_2; \alpha_2)}_{L_{FZ0}(\alpha_2)}, \quad (2.2.16)$$

where the $L_{FZ0}(Y, v_i, e_i; \alpha_i)$ is the FZ0 loss function for α_i as given by Patton et al. (2019) in their Eq.(2.2.4). By minimizing the expectation of $\mathfrak{L}_{FZ0}(Y, v_1, v_2, e_1, e_2; \alpha_1, \alpha_2)$, the model parameters are estimated via:

$$\hat{\theta}_T = \arg\min_{\theta} \frac{1}{T} \sum_{t=1}^{T} \mathfrak{L}_{FZ0}(Y_t, v_{1,t}, v_{2,t}, e_{1,t}, e_{2,t}; \alpha_1, \alpha_2), \quad (2.2.17)$$

where $v_{i,t}$ and $e_{i,t}$ are the VaR and ES forecasts at time t, obtained with the information set available at time $t-1$, at two risk levels α_i, $i = 1$ and 2. Before evaluating the performance of the augmented models, an essential consideration is the selection of the hyper-parameter, the auxiliary level α_2. To find an optimized α_2, we propose to use time series cross-validation, which is discussed in the following section.

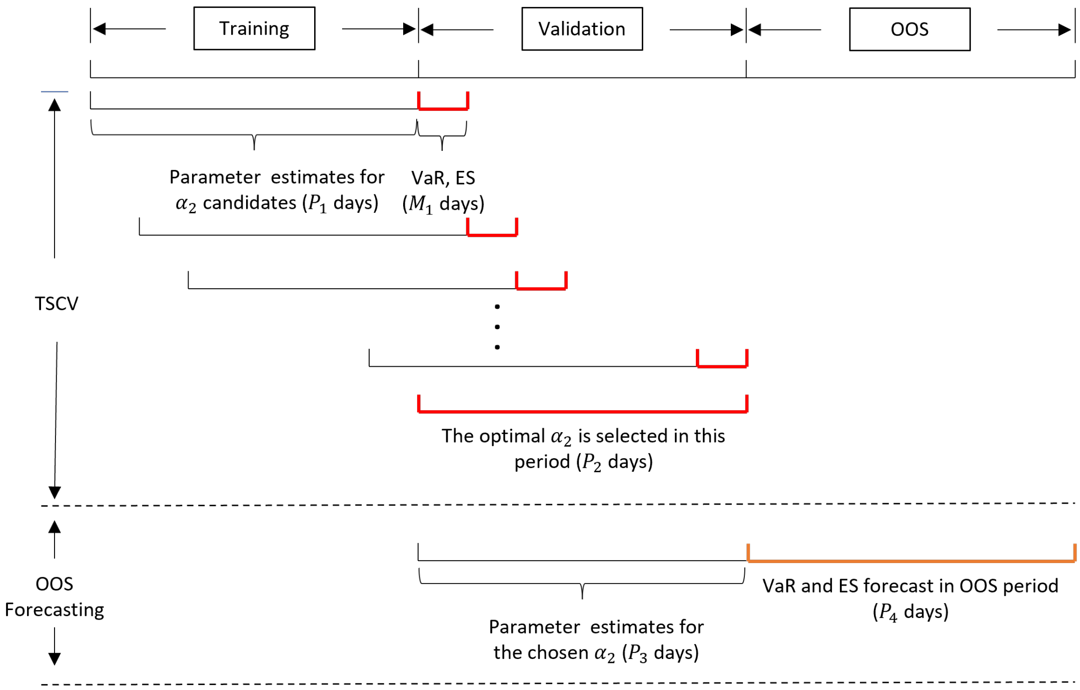
Time series cross-validation

For the above augmented models, α_2 is the hyper-parameter to be determined. Cross-validation has been introduced as a method to help choose the best hyper-parameters for models in general (see, for example, Hart (1994) for a description of this methodology). However, for time series, this method cannot be used in its classic form. Thus, we apply a special version of cross-validation that is suitable for time series applications, proposed by Hyndman and Athanasopoulos (2018). Within this procedure, a series of validation sets are formed, each consisting of an equal-weighted segment of the time series observations. The corresponding training set consists of observations that occurred before the validation set. Therefore, no future information is used when making forecasts for the validation sets. One small change that we make is to use the rolling-windows, rather than the expanding-windows as proposed by Hyndman and Athanasopoulos (2018). This is due to the consideration of possible structural breaks, which could lead to large forecasting errors and result in the model being unreliable for forecasting. Since we use a large number of observations, 3000 and 7 years of observations as the original training set in our simulation and empirical study, this is a matter not to be ignored. Our TSCV procedure is illustrated in Figure 2.2.2.

2.3 Simulation study

In this section, we investigate the performance of the A-GAS models and compare it with that of the GAS models of Patton et al. (2019) via Monte Carlo simulations. To measure model performance, we use loss values and a range of

Figure 2.2.2: TSCV procedure for selecting the auxiliary level



Note: This figure presents the TSCV procedure for selecting the best auxiliary level α_2 for the A-GAS model.[6] Also, it illustrates the forecasting procedure for VaR and ES at the extreme level α_1 in the OOS period. The parameters estimated from the training set (P_1 days) are used to forecast the VaR and ES in the first M_1 days in the validation period (P_2 days). Then, the parameters are re-estimated using information from day $1 + M_1$ to day $P_1 + M_1$ to generate the VaR and ES for the following M_1 days. After $\lfloor P_2/M_1 \rfloor$ repetitions, VaR and ES are forecasted for the validation period. The α_2 value with the lowest average FZ0 loss value over the validation period is selected as the optimal α_2 that is used in the OOS period. To obtain VaR and ES forecasts over the OOS period (P_4 days), the parameters of the A-GAS model with the optimal α_2 value are estimated using data from the last P_3 days prior to the OOS period. For the length of the time interval of P_3, the upper bound is the last day before the OOS period, and the lower bound can be chosen arbitrarily before the upper bound.

backtests. In the following we consider an extreme significance level as $\alpha_1 = 0.1\%$. We choose the data generation process (DGP) as the GJR-GARCH (1,1) model with skewed t distribution (we label it as GJR-GARCH-SKT) which considers the leverage effect, and it is among the most suitable model for volatility and VaR forecasting (Liu and Hung, 2010). Specifically, the DGP is:

$$\begin{aligned} Y_t &= \sigma_t \eta_t,\ \eta_t \sim iid\ F_\eta(0,1), \\ \sigma_t^2 &= \omega + \gamma Y_{t-1}^2 + \delta \mathbf{1}\{Y_{t-1} < 0\} Y_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned} \tag{2.3.1}$$

where the parameter values of the DGP are set to be $(\omega,\ \gamma,\ \delta,\ \beta) = (0.0225,$ $0.0065,\ 0.1779,\ 0.8835)$, and the error term η_t follows a skewed t distribution of Hansen (1994) with degrees of freedom $\nu = 7.5269$ and skewness parameter $\lambda = -0.1455$.[7]

In the specification of the A-GAS model, we consider a variety of values for α_2, specifically {2.5%, 5%,7.5%, 10%, 12.5%, 15%, 17.5%, 20%} is the set of possible values for the auxiliary significance level.[8] The simulation is based on 1000 replications. The whole sample size $T = 9000$ is divided into three equal segments, specifically training period, validation period and the OOS period. The first 3000 days of the sample is the training set used as a fixed window to obtain the parameter estimates for VaR and ES forecasts in the next 3000 days (which is the validation period for the selection of α_2). For the last 3000 days, the OOS period, we produce forecasts of risk measures for evaluation. TSCV selects the optimum α_2 obtained via minimizing the average loss in the validation period in each replication, and this α_2 will be used over the entire OOS period. Therefore, the A-GAS model with TSCV provides the lowest average loss over the validation period.

Table 2.3.1 presents the FZ0 loss values and loss reductions of the A-GAS models at the extreme level $\alpha_1 = 0.1\%$ for different α_2 values.[9] If α_2 is chosen by TSCV, the two A-GAS models outperform the classical alternative models. The loss reduction is defined as the relative reduction in the loss value of the risk measures obtained by the augmented models compared to their corresponding GAS model. The loss reduction of the augmented GAS one-factor model and the hybrid model are approximately 26% and 16%. The A-GAS models with TSCV are found to be the best GAS-type models, with loss values of 1.702 and 1.719 for the A-GAS-1F and the A-Hybrid model, equivalent to a 28.7% and 19.4% loss reduction, respectively. We further compare the loss obtained by the augmented models with the loss value from the "true" VaR and ES calculated from the DGP (when no model risk is present, labeled as GJR-G-True), as well as with the loss values obtained by estimating the DGP model (obtained when only parameter estimation risk is present, labeled as GJR-G-Est). The results suggest that the augmented models estimated via TSCV lead to risk values that have losses very close to the true loss values, complimenting the accuracy of the risk forecasts.

Next, we compare the augmented GAS models with the alternatives in terms of backtests of the risk forecasts. Three backtests are considered. First, we implement the unconditional coverage (UC) test proposed by Kupiec (1995) which uses the proportion of failures as its main tool to evaluate VaR. Second, the conditional coverage (CC) test proposed by Christoffersen (1998) is considered, and this test addresses the clustering of failures. Third, to evaluate the ES forecasts, we employ the bootstrap (BS) test of McNeil and Frey (2000), which focuses on

Table 2.3.1: Average loss values of the A-GAS models in simulation study

	A-GAS-1F			A-Hybrid		
α_2	Validation loss	OOS Forecast loss	Loss reduction	Validation loss	OOS Forecast loss	Loss reduction
2.5%	1.813	1.805	24.4%	1.813	1.796	15.8%
5%	1.783	1.773	25.7%	1.781	1.778	16.7%
7.5%	1.772	1.765	26.1%	1.794	1.773	16.9%
10%	1.802	1.763	26.2%	1.793	1.773	16.9%
12.5%	1.778	1.767	26.0%	1.803	1.778	16.6%
15%	1.767	1.769	25.9%	1.813	1.784	16.4%
17.5%	1.812	1.766	26.1%	1.820	1.789	16.2%
20%	1.786	1.762	26.2%	1.805	1.792	16.0%
TSCV	1.618	1.702	28.7%	1.644	1.719	19.4%
GAS	-	2.388	-	-	2.134	-
GJR-G-Est	-	1.691	-	-	1.691	-
GJR-G-True	-	1.675	-	-	1.675	-

Note: This table presents the average loss values over the validation and OOS forecasting period obtained from 1000 replications of VaR and ES estimations at the level $\alpha_1 = 0.1\%$, assuming a GJR-GARCH-SKT as the DGP. The full sample size is $T = 9000$, with 3000 as the training sample, 3000 as the validation sample, and 3000 as the OOS sample. The first 8 rows correspond to the A-GAS models with different α_2 values, whilst row 9 corresponds to the A-GAS models with TSCV. Row 10 corresponds to the GAS model of Patton et al. (2019). The last two rows are for the GJR-GARCH-SKT, where the GJR-G-Est corresponds to the GJR-GARCH-SKT model with the parameters re-estimated in every simulation path, and GJR-G-True corresponds to the GJR-GARCH-SKT model with the true parameters of the DGP. The loss values of the one-factor models are located in the left panel and loss values for the hybrid models are in the right panel. The columns "Validation loss" and "OOS forecast loss" present the average loss value of the augmented GAS models during the validation period and the OOS forecasting period, based on the FZ0 loss function. The Column "Loss reduction" presents the relative reduction in the loss value of the risk measures obtained by the augmented models compared to their corresponding GAS model.

the discrepancies between the observed returns and the ES forecasts for the periods in which the return exceeds the VaR forecast. We calculate the rate that the null is rejected at 5% significance level, and we call this the Rejection Rate. This is reported over the OOS period.[10]

Table 2.3.2 presents the backtest rejection rate at the extreme level $\alpha_1 = 0.1\%$ for VaR and ES of the two A-GAS models, the GAS model, the true model with true parameter labeled as GJR-G-Ture, and the true model with estimated parameter labeled as GJR-G-Est. In general, the A-GAS models outperform the GAS model. The A-GAS models with TSCV provide the lowest backtest rejection rates except for the BS backtest results of the A-Hybrid model. On the other hand, the A-GAS-1F model with TSCV has the best performance in the BS backtest. The GJR-GARCH-SKT model performs best in terms of backtest rejection rate, which is as expected, because this is the DGP model used for the simulation. It is important to note that using the true DGP model is only possible in a simulation setup, whilst in practice the true DGP is unknown.

Additionally, we explore model performance for different values of α_1. Table 2.3.3 presents the relative loss reduction obtained by the A-GAS models compared to the GAS models, for different extreme levels of α_1. We consider augmented models for various values of α_1 and α_2. Specifically, $\alpha_1 \in \{0.1\%, 0.25\%, 0.5\%, 0.75\%, 1\%, 2.5\%\}$ and $\alpha_2 \in \{2.5\%, 5\%, 7.5\%, 10\%, 12.5\%, 15\%, 17.5\%, 20\%\}$. We set the restriction that α_1 must be lower than α_2. As α_1 decreases, the loss reduction obtained via the A-GAS model as compared to the corresponding GAS model increases for all α_2 candidates. Also, we find that the A-GAS models with TSCV have the greatest improvement in terms of loss values. The greatest

Table 2.3.2: Backtest rejection rates in simulation study

	A-GAS-1F			A-Hybrid		
α_2	UC	CC	BS	UC	CC	BS
2.5%	35.8%	22.7%	29.9%	37.4%	25.6%	25.4%
5%	36.1%	23.6%	28.7%	38.0%	25.9%	28.5%
7.5%	35.8%	23.0%	29.6%	36.5%	26.1%	30.6%
10%	35.7%	23.6%	29.8%	37.8%	26.6%	29.2%
12.5%	35.0%	23.6%	28.9%	37.6%	28.1%	31.1%
15%	36.6%	24.6%	29.5%	37.2%	26.7%	32.6%
17.5%	36.1%	23.7%	28.2%	37.5%	27.4%	33.3%
20%	34.4%	22.0%	27.0%	37.1%	26.7%	32.5%
TSCV	27.9%	15.7%	22.9%	31.5%	21.9%	30.3%
GAS	47.8%	37.4%	40.0%	47.9%	38.3%	34.0%
GJR-G-Est	13.2%	7.7%	20.9%	13.2%	7.7%	20.9%
GJR-G-True	8.0%	5.7%	21.0%	8.0%	5.7%	21.0%

Note: This table presents the backtest rejection rates obtained from 1000 replications, indicating the frequency of backtest rejections at 5% significance level. The DGP used in the simulation is the GJR-GARCH-SKT, $T = 9000$ and the risk level used to compute VaR and ES is $\alpha_1 = 0.1\%$. The first 8 rows correspond to the A-GAS models with different α_2 values, whilst row 9 corresponds to the A-GAS model with TSCV. Columns 2-4 and columns 5-7 present the rejection rates for the Unconditional Coverage (UC), Conditional Coverage (CC) and Bootstrapping backtest (BS), respectively.

reduction is obtained for $\alpha_1 = 0.1\%$ and it is about 15% and 20% in relative terms for the A-GAS-1F and A-Hybrid models, respectively.

Overall, based on the results of the above simulation studies, incorporating information from an auxiliary level of significance improves the performance of the GAS models when α_1 is extremely small.[11] Also, TSCV is shown to be highly effective to choose the hyper-parameter α_2.

2.4 Empirical study

2.4.1 Data description

To evaluate the empirical forecast performance of the proposed models, we study daily returns from four oil futures, the WTI crude oil, Brent crude oil, Gas oil

Table 2.3.3: Relative loss reduction of A-GAS models for various α_1 levels

	A-GAS-1F						A-Hybrid					
$\alpha_2 \backslash \alpha_1$	0.1%	0.25%	0.5%	0.75%	1%	2.5%	0.1%	0.25%	0.5%	0.75%	1%	2.5%
2.5%	24.39%	16.60%	11.88%	8.89%	6.55%	N/A	15.81%	13.54%	7.50%	4.39%	3.11%	N/A
5%	25.73%	18.65%	14.00%	11.03%	9.12%	2.61%	16.68%	14.94%	8.73%	5.95%	4.66%	1.54%
7.5%	26.07%	19.35%	14.87%	11.84%	9.91%	3.70%	16.90%	15.29%	9.45%	6.66%	5.22%	2.12%
10%	26.15%	19.74%	15.17%	12.39%	10.34%	4.16%	16.91%	15.34%	9.81%	7.04%	5.71%	2.51%
12.5%	25.99%	20.20%	15.55%	12.65%	10.54%	4.39%	16.65%	15.70%	9.92%	7.12%	5.85%	2.72%
15%	25.90%	19.88%	15.53%	12.66%	10.60%	**4.52%**	16.38%	15.68%	9.99%	7.30%	5.97%	2.78%
17.5%	26.05%	20.15%	15.51%	12.68%	10.54%	4.49%	16.17%	15.61%	9.94%	7.22%	5.97%	2.78%
20%	26.22%	20.01%	15.56%	12.65%	10.52%	4.43%	16.03%	15.51%	10.05%	7.16%	5.88%	**2.79%**
TSCV	**28.70%**	**20.58%**	**15.76%**	**12.73%**	**10.65%**	4.43%	**19.42%**	**16.67%**	**10.39%**	**7.44%**	**6.10%**	2.75%

Note: This table presents the relative loss reduction of the A-GAS models as compared to their corresponding GAS models, based on 1000 replications of VaR and ES estimations for a GJR-GARCH-SKT DGP with $T = 9000$. Loss reduction values for the one-factor GAS models are on the left and for the Hybrid GAS/GARCH model are on the right. The first column presents the α_2 candidates from 2.5% up to 20% and the last row reports the values obtained by applying TSCV in the selection of α_2. The rest of the columns present the relative loss reduction for different target levels (α_1). In each column, the α_2 with the highest reduction is highlighted in bold. When $\alpha_1 \geq \alpha_2$, the loss values are reported as N/A.

(GO) and Heating oil (HO).[12] We choose these series because they are typically characterized by high volatility. The sample period is between 1 January 2000 and 8 January 2021, and our data source is DataStream.[13]

Table 2.4.1 presents the summary statistics of these four series over the full sample period. All return series exhibit substantial kurtosis of between 8 and 20. The table also shows the sample VaR and ES for four significance levels: 0.1%, 0.25%, 0.5% and 1%, and Panel B presents the estimated parameters of the GARCH (1,1) model with a skewed t distribution, fitted to the de-meaned returns, with the model defined as:

$$\begin{aligned} Y_t &= \sigma_t \eta_t, \ \eta_t \sim iid \text{ Skew } t(0, 1, \nu, \lambda), \\ \sigma_t^2 &= \omega + \beta \sigma_{t-1}^2 + \gamma Y_{t-1}^2. \end{aligned} \tag{2.4.1}$$

The full sample is divided into a training period (January 2000 to December 2006), a validation period (January 2007 to December 2013), and an OOS forecasting period (January 2014 to January 2021). In the validation period, we employ the TSCV introduced in Section 2.2.3 with eight candidates of $\alpha_2 \in \{2.5\%, 5\%, 7.5\%, 10\%, 12.5\%, 15\%, 17.5\%, 20\%\}$, and the α_2 value which provides the lowest loss value is selected as the optimal α_2 for forecasting OOS. Then we produce risk forecasts for $\alpha_1 = 0.1\%$ for the OOS period, and the forecasting performance is evaluated in the OOS period.

Table 2.4.1: Summary statistics and parameter estimates for oil futures

	WTI	Brent	GO	HO
Panel A: Summary statsitics				
Mean (annualized)	6.301	3.765	3.445	3.959
Std. dev. (annualized)	42.077	36.969	34.178	36.895
Skewness	0.037	-0.590*	-0.254*	-0.653*
Kurtosis	20.406*	14.411*	8.017*	10.005*
VaR ($\alpha = 0.1\%$)	-12.970	-11.066	-11.483	-15.758
VaR ($\alpha = 0.25\%$)	-10.806	-9.688	-8.610	-9.225
VaR ($\alpha = 0.5\%$)	-9.109	-7.562	-6.757	-8.173
VaR ($\alpha = 1\%$)	-7.089	-6.392	-5.633	-6.217
ES ($\alpha = 0.1\%$)	-22.800	-19.257	-14.882	-19.487
ES ($\alpha = 0.25\%$)	-15.974	-13.685	-11.724	-14.663
ES ($\alpha = 0.5\%$)	-12.952	-10.986	-9.741	-11.652
ES ($\alpha = 1\%$)	-10.460	-8.982	-7.851	-9.360
Panel B: Parameter estimates				
ω	0.064*	0.036*	0.014*	0.029*
β	0.914*	0.926*	0.951*	0.936*
γ	0.076*	0.069*	0.047*	0.060*
ν	7.583*	6.985*	8.098*	7.029*
λ	-0.094*	-0.075*	-0.058*	-0.030*

Note: Summary statistics and parameter estimates for the four futures return series, over the full sample period from January 2000 to January 2021. Panel A reports the annualized mean and standard deviation of the returns expressed in percentages, the skewness, kurtosis, as well as the sample VaR and ES estimates for four different values of risk level α. * denotes values of skewness (kurtosis) significantly different from zero (3) at 5% level. Panel B presents the estimated parameters of the GARCH (1,1) model with skewed t distributed errors, with a * symbolizing parameter values that are significant at 5% level.

2.4.2 Estimation results

Table 2.4.2 presents the loss values for VaR and ES at the extreme level $\alpha_1 = 0.1\%$ for different values of α_2 for the four series considered.[14] The optimal value of α_2 is found in the range from 10% to 15% for the A-GAS-1F model, while for the A-Hybrid model we find that α_2 is above 15%, except for $\alpha_2 = 7.5\%$ for Brent. Table 2.4.3 presents the estimated parameters together with their standard errors for the GAS and A-GAS models with α_2 chosen by TSCV over the validation period. It can be noted that the loss values of the A-GAS-1F model are below the loss values of the GAS-1F for all series considered. The A-Hybrid model has losses below those of the Hybrid model for WTI and Brent. Both A-GAS models estimated on all four energy commodity futures demonstrate a higher value of γ compared to the original GAS models. This implies that the estimated VaR and ES of the A-GAS models at the extreme level α_1 are influenced by the value of VaR and ES at the auxiliary level α_2. A higher value of γ in the A-GAS models indicates that the κ_t process (and the VaR and ES processes) is more reactive to the forcing variable s_{t-1}. The estimates for the γ of original GAS models are not statistically significant in all four oil futures, while the A-GAS models perform better in the Brent, GO, and HO in terms of the significance of the parameter γ.

2.4.3 Out-of-sample results

We now turn to the OOS forecast performance of the A-GAS models at the extreme level $\alpha_1 = 0.1\%$, as compared to a total of fourteen alternative models. Six non-parametric models are considered as benchmarks, including the traditional

Table 2.4.2: Average losses during the validation period for various α_2 levels

	A-GAS-1F				A-Hybrid			
α_2	WTI	Brent	GO	HO	WTI	Brent	GO	HO
2.5%	2.422	2.486	2.091	2.520	2.494	2.622	1.915	2.667
5%	2.379	2.632	2.196	2.531	2.523	2.425	1.925	2.641
7.5%	2.505	2.509	2.044	2.534	2.493	**2.350**	2.040	2.580
10%	2.293	2.578	2.101	**2.288**	2.502	2.467	1.969	2.466
12.5%	2.286	**2.414**	**1.885**	2.440	2.399	2.414	1.919	2.485
15%	**2.266**	2.463	1.973	2.342	2.465	2.420	**1.895**	2.427
17.5%	2.268	2.524	1.975	2.425	**2.343**	2.476	1.981	2.470
20%	2.285	2.436	1.991	2.452	2.437	2.527	2.037	**2.424**

Note: This table presents the average loss values of VaR and ES at the extreme level $\alpha_1 = 0.1\%$ in the validation period for eight α_2 values, estimated for the return series of four oil futures from January 2007 to December 2013. The left panel indicates the average FZ0 loss value of Eq. (2.2.4) for the A-GAS-1F model whilst the values for the A-Hybrid model are presented in the right panel. The lowest value in each column is in bold, and the corresponding α_2 is selected for the OOS period.

rolling window methods with window lengths of 500, 1000 and 1500 trading days and rolling window methods based on the Cornish-Fisher expansion (Cornish and Fisher, 1938), with the same window lengths as the first three models. Four prevailing GARCH models are also considered as benchmarks, namely, the GARCH model with normal distribution (GARCH-N), GARCH with skewed t distribution (GARCH-SKT), GARCH with empirical distribution function (GARCH-EDF) in which case VaR and ES are estimated from the sample VaR and ES of the estimated standardized residuals obtained from the GARCH model, and the GJR-GARCH with skewed t distribution (GJR-GARCH-SKT). We next consider four models introduced by Patton et al. (2019): the two-factor GAS model (GAS-2F), the one-factor GAS model (GAS-1F), the GARCH model using FZ loss minimization (GARCH-FZ), and the hybrid GAS/GARCH model (Hybrid). Finally, we consider the two proposed augmented GAS models, the A-GAS-1F model and

Table 2.4.3: Parameter estimates of the GAS and A-GAS models

	Panel A: WTI				Panel B: Brent			
	GAS-1F	Hybrid	A-GAS-1F	A-Hybrid	GAS-1F	Hybrid	A-GAS-1F	A-Hybrid
β	0.986	0.810	0.971	0.870	0.942	0.785	0.955	0.927
	(0.004)	(0.015)	(0.028)	(0.019)	(0.03)	(0.020)	(0.005)	(0.026)
γ	0.002	0.000	0.068	0.077	0.002	0.000	0.051	0.056
	(0.015)	(0.000)	(0.168)	(0.237)	(0.007)	(0.000)	(0.096)	(0.015)
δ	-	0.058	-	0.064	-	0.074	-	0.039
	-	(0.009)	-	(0.009)	-	(0.038)	-	(0.013)
a_1	-6.271	-8.950	-5.198	-7.443	-9.689	-8.246	-7.082	-9.393
	(6.031)	(0.621)	(5.340)	(1.910)	(1.360)	(1.299)	(1.689)	(0.159)
b_1	-6.727	-9.745	-5.711	-7.444	-9.935	-8.362	-8.911	-14.490
	(4.485)	(0.580)	(5.354)	(1.881)	(1.372)	(1.819)	(1.671)	(0.127)
a_2		-	-1.191	-1.596	-	-	-1.952	-3.662
	-	-	(14.142)	(5.662)	-	-	(6.905)	(2.001)
b_2	-	-	-2.388	-2.991	-	-	-3.237	-7.8646
	-	-	(14.268)	(5.533)	-	-	(6.866)	(3.097)
Ave. loss	2.274	2.287	2.058	2.057	2.276	2.199	2.152	2.127

	Panel C: GO				Panel D: HO			
	GAS-1F	Hybrid	A-GAS-1F	A-Hybrid	GAS-1F	Hybrid	A-GAS-1F	A-Hybrid
β	0.963	0.903	0.983	0.830	0.989	0.999	0.981	0.957
	(0.003)	(0.002)	(0.001)	(0.030)	(0.002)	(0.000)	(0.042)	(0.105)
γ	0.005	0.001	0.013	0.014	0.000	0.000	0.034	0.079
	(0.004)	(0.000)	(0.004)	(0.066)	(0.003)	(0.000)	(0.008)	(0.103)
δ	-	0.068	-	0.100	-	0.004	-	0.038
	-	(0.000)	-	(0.002)	-	(0.000)	-	(0.019)
a_1	-7.004	-5.878	-6.711	-5.943	-8.739	-7.823	-7.165	-7.823
	(0.731)	(0.743)	(0.221)	(0.326)	(2.156)	(2.126)	(0.243)	(2.550)
b_1	-8.660	-6.555	-6.737	-7.409	-8.966	-7.823	-10.360	-7.916
	(0.849)	(0.68)	(0.217)	(0.353)	(2.314)	(2.149)	(0.138)	(3.154)
a_2	-	-	-2.171	-1.773	-	-	-2.891	-1.403
	-	-	(0.644)	(1.305)	-	-	(0.410)	(8.840)
b_2	-	-	-2.966	-3.261	-	-	-4.231	-2.798
	-	-	(0.684)	(1.299)	-	-	(0.777)	(10.033)
Ave. loss	1.907	1.676	1.828	1.732	2.192	2.015	2.072	2.034

Note: This table presents parameter estimates and standard errors (in parentheses) for two GAS models and two A-GAS models used to forecast VaR and ES at the extreme level $\alpha_1 = 0.1\%$ for four oil futures series over the in-sample period from January 2007 to December 2013. For each return series, the first two columns in each panel present the parameter estimates for the one-factor GAS model and the Hybrid GAS/GARCH model, and the following two columns indicate the parameter values for the one-factor A-GAS model and the A-Hybrid GAS/GARCH model, respectively. The last row of each panel presents the average in-sample FZ0 loss for the four return series.

the A-hybrid model. We estimate the parametric and semiparametric models using the first 7 years starting with 2007 as the in-sample period, and retain the parameter estimates to build forecasts for the OOS period.

Table 2.4.4 presents the p-values of the VaR and ES backtests in the OOS period (from January 2014 to January 2021) and over the COVID-19 period (from January 2020 to January 2021) for 16 models and for four oil futures. As before, we forecast risk at the significance level of $\alpha_1 = 0.1\%$. We find that, over the whole OOS period, the augmented models can pass the VaR backtests (UC and CC) for the Brent series. Considering the ES backtest (BS), our models provide reasonable backtest results for the time series of WTI, GO, and HO. During the COVID-19 period, a time marked by exceptional market volatility and unpredictability, the models proposed by Patton et al. (2019) faced challenges in passing the Unconditional Coverage (UC) test. In contrast, our models, especially the A-Hybrid model, demonstrated a significant improvement in performance under these testing conditions. This is particularly evident in the VaR backtests.

To provide a more robust evaluation, we also employ a comparison based on the FZ0 loss function to assess the performance of the models considered. Table 2.4.5 presents the average losses and the ranks of the models based on loss values at the level $\alpha_1 = 0.1\%$ over the OOS period and the COVID-19 period. Over the whole OOS period, the A-GAS-1F model provides the lowest average loss for Brent and HO. Even though the A-GAS-1F model is not the best-performing model for all four futures, its average ranking is the best among all the models. During the highly volatile COVID-19 period, characterized by a surge in extreme losses, the A-Hybrid model outperformed other models in terms

Table 2.4.4: Out-of-sample backtest performance for oil futures

	UC test (VaR) p-values				CC test (VaR) p-values				BS test (ES) p-values			
Panel A: Jan 2014 to Jan 2021												
	WTI	Brent	GO	HO	WTI	Brent	GO	HO	WTI	Brent	GO	HO
RW-500	0.012	0.001	0.010	0.003	0.002	0.000	0.002	0.001	0.032	0.000	0.008	0.030
RW-1000	0.012	**0.051**	0.040	0.013	0.002	0.004	0.004	0.002	**0.108**	**0.138**	0.014	0.048
RW-1500	**0.148**	0.014	0.000	**0.150**	0.007	0.002	0.000	**0.352**	**0.110**	**0.374**	**0.104**	**0.102**
CF-500	**0.396**	**0.160**	**0.366**	**0.399**	**0.694**	**0.369**	0.007	**0.697**	**0.284**	**0.570**	**0.288**	**0.410**
CF-1000	**0.859**	**0.512**	**0.132**	**0.399**	**0.982**	**0.806**	**0.319**	**0.697**	**0.490**	0.000	**0.794**	**0.292**
CF-1500	**0.532**	**0.512**	**0.366**	**0.529**	**0.822**	**0.806**	**0.661**	**0.820**	0.000	0.000	**0.338**	0.000
GARCH-N	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.008
GARCH-SKT	0.000	0.001	**0.132**	0.003	0.000	0.000	0.006	0.012	**0.236**	**0.516**	**0.140**	**0.054**
GARCH-EDF	0.000	0.000	0.010	0.000	0.000	0.000	0.002	0.000	**0.498**	**0.694**	**0.196**	**0.740**
GJR-GARCH-SKT	0.003	0.003	**0.366**	0.001	0.001	0.001	0.007	0.003	**0.488**	**0.606**	**0.492**	**0.098**
GARCH-FZ	0.000	0.000	0.010	0.003	0.000	0.000	0.037	0.012	**0.294**	**0.674**	**0.146**	0.012
GAS-1F	**0.396**	**0.887**	**0.366**	0.047	0.007	**0.988**	**0.661**	**0.137**	**0.472**	**0.474**	**0.784**	**0.146**
Hybrid	0.046	**0.051**	**0.132**	0.001	**0.136**	0.004	**0.319**	0.000	**0.140**	**0.146**	**0.078**	0.006
GAS-2F	**0.148**	**0.417**	**0.132**	0.047	0.007	0.007	**0.319**	**0.137**	**0.368**	**0.594**	**0.706**	**0.084**
A-GAS-1F	0.000	**0.051**	0.040	0.013	0.000	0.004	0.004	0.044	**0.062**	0.028	0.044	**0.550**
A-Hybrid	0.001	**0.051**	0.000	0.013	0.003	**0.147**	0.000	0.044	**0.810**	0.046	**0.952**	0.028
Panel B: Jan 2020 to Jan 2021												
	UC test (VaR) p-values				CC test (VaR) p-values				BS test (ES) p-values			
	WTI	Brent	GO	HO	WTI	Brent	GO	HO	WTI	Brent	GO	HO
RW-500	0.029	0.003	0.002	0.002	0.002	0.001	0.001	0.001	**0.526**	0.078	**0.074**	**0.094**
RW-1000	0.000	0.003	0.002	0.002	0.000	0.001	0.001	0.001	**0.094**	0.072	**0.596**	**0.080**
RW-1500	0.000	0.000	0.000	0.030	0.000	0.000	0.000	**0.093**	**0.118**	**0.420**	**0.108**	**0.482**
CF-500	**0.262**	**0.275**	0.028	**0.268**	**0.531**	**0.548**	0.002	**0.539**	0.000	0.000	**0.538**	0.000
CF-1000	**0.262**	**0.275**	0.028	**0.268**	**0.531**	**0.548**	**0.089**	**0.539**	0.000	0.000	**0.478**	0.000
CF-1500	**0.262**	**0.275**	0.028	**0.472**	**0.531**	**0.548**	**0.089**	0.000	0.000	0.000	**0.508**	0.000
GARCH-N	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.001	**0.076**	**0.052**	**0.138**	**0.074**
GARCH-SKT	0.002	0.000	0.002	0.030	0.001	0.000	0.001	**0.093**	**0.506**	**0.528**	**0.276**	**0.498**
GARCH-EDF	0.000	0.000	0.002	0.002	0.000	0.000	0.001	0.001	**0.530**	**0.592**	**0.088**	**0.484**
GJR-GARCH-SKT	0.029	0.003	0.028	0.030	0.002	0.001	0.002	**0.093**	**0.456**	**0.544**	**0.552**	**0.556**
GARCH-FZ	0.000	0.000	0.028	0.002	0.000	0.000	**0.089**	0.009	**0.184**	**0.556**	**0.494**	**0.402**
GAS-1F	0.029	0.031	**0.260**	0.030	0.002	**0.097**	**0.528**	**0.093**	**0.508**	**0.520**	0.000	**0.480**
Hybrid	0.002	0.003	0.028	0.002	0.009	0.001	**0.089**	0.001	**0.070**	**0.082**	**0.498**	**0.092**
GAS-2F	0.002	0.003	0.028	0.030	0.001	0.001	**0.089**	**0.093**	**0.322**	**0.570**	**0.504**	**0.492**
A-GAS-1F	0.002	0.003	0.002	0.030	0.009	0.001	0.001	**0.093**	**0.516**	**0.604**	**0.156**	**0.488**
A-Hybrid	0.029	**0.275**	0.002	**0.472**	**0.090**	**0.548**	0.001	0.000	**0.510**	0.000	**0.362**	0.000

Note: This table presents the p-values of two VaR backtests and an ES backtest for four oil futures, over the whole OOS period (Panel A) and the COVID-19 period (Panel B) for 16 risk forecasting models at level $\alpha_1 = 0.1\%$. Columns 2-5 and 6-9 present the results for the Unconditional Coverage (UC) and the Conditional Coverage (CC) backtest for the evaluation of VaR. The last 4 Columns present the results of the Bootstrapping (BS) backtest for the evaluation of ES. Values greater than 0.05 (indicating no evidence against optimality at 5% significance level) are in bold.

of average loss for all the oil futures, except for the GO series. The A-GAS-1F model provides stable performance during the COVID-19 period, achieving the best average ranking overall and demonstrating strong results. Both models showcase their out-performance in forecasting market risks in navigating extreme market conditions.

While average losses are a useful tool to consider forecast performance out-the-sample, they do not provide information on the significance of the loss differences between models. Figure 2.4.1 presents the results of the Diebold–Mariano (DM) test that performs pairwise model comparisons based on loss differences over the OOS period, with the null hypothesis that the row model and the column model have equal loss values. The A-GAS-1F model has superior performance for all series, especially for Brent and HO, outperforming all alternative models considered.[15]

2.5 Robustness check

In practice, time series are often characterized by the presence of structural breaks in the fitted models. As such, we perform a robustness check by estimating the parameters of all models using rolling windows. Also, the forecasts of VaR and ES are generated via rolling windows, and each model is re-estimated every 30 trading days using a window length of 1,805 observations (7 years) starting from January 2007. The remaining period until January 2021 (1,805 days in total) is the OOS period used to evaluate one-day-ahead VaR and ES estimates.[16] We find that the results are similar to the ones reported in earlier sections of this paper.

Table 2.4.5: Out-of-sample losses and loss rankings for oil futures

	Average loss				Loss ranking				
Panel A: Jan 2014 to Jan 2021									
	WTI	Brent	GO	HO	WTI	Brent	GO	HO	Average
RW-500	4.315	4.056	3.869	5.031	15	14	16	15	15
RW-1000	3.908	3.803	3.197	4.569	11	13	13	12	12.25
RW-1500	3.911	4.116	3.314	3.514	12	15	14	5	11.5
CF-500	3.599	2.994	2.998	3.474	9	6	11	3	7.25
CF-1000	3.272	3.035	2.781	3.870	6	7	9	11	8.25
CF-1500	3.310	3.175	2.730	3.575	7	9	8	7	7.75
GARCH-N	4.912	4.770	3.526	5.131	16	16	15	16	15.75
GARCH-SKT	3.335	2.990	2.631	3.703	8	5	6	8	6.75
GARCH-EDF	3.688	3.425	2.836	3.804	10	10	10	10	10
GJR-GARCH-SKT	*2.886*	2.736	**2.333**	3.558	2	3	1	6	*3*
GARCH-FZ	4.127	3.498	2.428	4.697	14	11	3	13	10.25
GAS-1F	**2.665**	2.890	2.472	3.509	1	4	4	4	3.25
Hybrid	3.986	3.147	2.568	4.970	13	8	5	14	10
GAS-2F	3.055	3.572	2.648	*3.448*	5	12	7	2	6.5
A-GAS-1F	2.997	**2.271**	*2.404*	**2.966**	4	1	2	1	**2**
A-Hybrid	2.911	*2.523*	3.061	3.717	3	2	12	9	6.5
Panel B: Jan 2020 to Jan 2021									
	Average loss				Loss ranking				
	WTI	Brent	GO	HO	WTI	Brent	GO	HO	Average
RW-500	10.824	11.153	9.277	9.962	11	12	16	14	13.25
RW-1000	12.517	12.562	7.458	11.196	14	14	13	16	14.25
RW-1500	13.743	14.715	8.654	5.689	16	16	15	6	13.25
CF-500	6.232	5.715	4.693	5.188	5	3	7	4	4.75
CF-1000	7.930	7.155	4.315	6.547	10	7	4	7	7
CF-1500	7.678	7.593	4.437	3.339	9	8	6	3	6.5
GARCH-N	11.561	12.780	8.324	9.890	12	15	14	13	13.5
GARCH-SKT	6.624	7.105	4.765	6.802	6	6	8	8	7
GARCH-EDF	7.639	8.416	5.939	7.011	8	9	12	9	9.5
GJR-GARCH-SKT	4.871	5.868	*3.372*	5.657	4	4	2	5	3.75
GARCH-FZ	12.183	9.066	**2.613**	10.237	13	11	1	15	10
GAS-1F	4.449	6.730	4.364	8.340	3	5	5	12	6.25
Hybrid	13.683	8.995	5.391	8.048	15	10	9	10	11
GAS-2F	7.115	11.617	5.898	8.265	7	13	11	11	10.5
A-GAS-1F	*3.699*	*3.065*	4.248	*2.887*	2	2	3	2	**2.25**
A-Hybrid	**2.713**	**2.925**	5.544	**2.330**	1	1	10	1	*3.25*

Note: This table presents the average losses and loss rankings (with the best performing model ranked 1 and the worst ranked 16) based on average FZ0 losses, for VaR and ES forecasts at level $\alpha_1 = 0.1\%$ of four oil futures, over the OOS period (Panel A) and the COVID-19 period (Panel B). Columns 2-5 present the average FZ0 losses, with the lowest (second lowest) in each column shown in bold (italics). Columns 6-9 present the loss rankings. The last column presents the average rank across the four series, with the best (second best) model shown in bold (italics).

Figure 2.4.1: Diebold-Mariano test based on a fixed window estimation

Note: This figure presents the Diebold–Mariano (DM) test results comparing the FZ0 losses over the OOS period from January 2014 to January 2021, for 16 models across four oil futures, comparing risk forecasts at level $\alpha_1 = 0.1\%$. Dark green (red) blocks mean that the row model has significantly lower (greater) average loss than the column model at 10% significance level; light green (yellow) blocks mean that the row model has lower (greater) average loss than the column model, but the difference is not significant.

Over the OOS period, the loss ranks of the A-GAS models slightly decrease, but the A-GAS-1F model has the best performance overall for most of the oil futures series considered. When considering the COVID-19 period in isolation, the A-GAS models, based on rolling window estimation, show the same superior performance as before.

2.6 Conclusion

This paper introduces augmented versions of the GAS models that jointly estimate risk at an extreme significance level and an auxiliary level of significance, with the purpose to improve on the forecasts of VaR and ES at an extreme level. By using TSCV to select the optimal auxiliary level, we document an improvement in the risk forecasts both in-sample and during the OOS periods considered. Our simulation study also highlights this improvement in terms of the forecast loss and the backtest rejection rates. We employ the proposed A-GAS models to forecast the VaR and ES of four oil futures over the period from January 2000 to January 2021. We compare these with forecasts made by fourteen alternative models, and we implement several backtests to compare their performance. The main finding is that VaR and ES forecasts obtained from the A-GAS models outperform the risk forecasts based on popular GARCH models or historical simulations, and they also lead to improved loss values compared with the original GAS models for three out of four future series considered. The A-GAS models perform even better during the COVID-19 period which is characterized by extreme losses. As such, the proposed augmented versions of popular GAS risk models can provide improved risk forecasts at extreme levels by utilizing the information from prevailing risk levels without considering exogenous information. Applications of these models to study the risk of other asset classes would be of future interest. Additionally, the proposed framework of estimating risk at extreme levels can be extended to more than two risk level or by considering alternative risk models.

Appendices

2.A Additional results

Table 2.A.1, presents backtest results for VaR and ES based on a rolling window estimation at 5% significance level, in the OOS period (from January 2014 to January 2021) and the COVID-19 period (from January 2020 to January 2021) for 16 models for $\alpha_1 = 0.1\%$. During the OOS period, the A-GAS models pass the VaR backtests for HO, and pass the ES backtest for all series. During the COVID-19 period, our models still pass the UC backtest for HO, the CC backtest for WTI and HO, and for WTI, Brent and GO they pass the BS bactkest. Similar to the fixed window estimation, backtesting VaR and ES individually might not be sufficient. Therefore, we investigate the loss score based on the FZ0 loss function for the models used in the rolling window estimation.

Table 2.A.2, presents the average FZ0 loss and the loss rankings at $\alpha_1 = 0.1\%$ for the whole OOS period (from January 2014 to January 2021) as well as for the COVID-19 period (from January 2020 to January 2021), for 16 models. Columns 2-5 and 6-9 present the average loss value and the ranks based on the loss for the four oil futures, respectively. The last column presents the average rank of the models across all return series. Looking at the whole OOS period, we observe that

Table 2.A.1: Out-of-sample backtest performance based on a rolling window estimation

	UC test (VaR) p-values				CC test (VaR) p-values				BS test (ES) p-values			
Panel A: Jan 2014 to Jan 2021												
	WTI	Brent	GO	HO	WTI	Brent	GO	HO	WTI	Brent	GO	HO
RW-500	0.012	0.001	0.010	0.003	0.002	0.000	0.002	0.001	0.048	0.014	0.010	0.028
RW-1000	0.012	**0.051**	0.040	0.013	0.002	0.004	0.004	0.002	**0.118**	**0.090**	0.020	0.038
RW-1500	**0.148**	0.014	0.000	**0.150**	0.007	0.002	0.000	**0.352**	**0.126**	**0.404**	**0.094**	**0.100**
CF-500	**0.396**	**0.160**	**0.366**	**0.399**	**0.694**	**0.369**	0.007	**0.697**	**0.298**	**0.602**	**0.342**	**0.350**
CF-1000	**0.859**	**0.512**	**0.132**	**0.399**	**0.982**	**0.806**	**0.319**	**0.697**	**0.548**	0.000	**0.728**	**0.328**
CF-1500	**0.532**	**0.512**	**0.366**	**0.529**	**0.822**	**0.806**	**0.661**	**0.820**	0.000	0.000	**0.348**	0.000
GARCH-N	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.000
GARCH-SKT	0.003	**0.160**	**0.366**	**0.150**	0.001	**0.369**	0.007	**0.352**	**0.576**	**0.572**	**0.560**	**0.068**
GARCH-EDF	**0.396**	0.014	**0.132**	**0.399**	**0.694**	0.002	0.006	**0.697**	**0.558**	**0.424**	**0.548**	**0.328**
GJR-GARCH-SKT	**0.148**	**0.417**	**0.563**	**0.150**	**0.348**	**0.716**	**0.846**	**0.352**	**0.418**	**0.530**	0.000	**0.066**
GARCH-FZ	0.003	0.014	**0.132**	0.003	0.001	0.048	0.006	0.012	**0.108**	**0.386**	**0.592**	0.042
GAS-1F	0.003	0.000	0.040	**0.150**	0.001	0.000	**0.120**	**0.352**	**0.100**	0.040	**0.308**	**0.100**
Hybrid	0.003	0.001	0.010	**0.399**	0.001	0.000	0.037	**0.697**	0.048	0.004	0.036	**0.378**
GAS-2F	0.000	0.001	0.000	0.003	0.000	0.000	0.000	0.012	0.022	**0.164**	**0.788**	**0.434**
A-GAS-1F	0.001	**0.160**	0.010	**0.864**	0.003	0.007	0.037	**0.983**	**0.288**	**0.462**	**0.182**	**0.534**
A-Hybrid	0.012	0.003	0.000	**0.150**	0.043	0.001	0.002	**0.352**	**0.072**	**0.154**	**0.792**	**0.150**
Panel B: Jan 2020 to Jan 2021												
	UC test (VaR) p-values				CC test (VaR) p-values				BS test (ES) p-values			
	WTI	Brent	GO	HO	WTI	Brent	GO	HO	WTI	Brent	GO	HO
RW-500	0.029	0.003	0.002	0.003	0.002	0.001	0.001	0.001	**0.510**	**0.090**	**0.078**	**0.072**
RW-1000	0.000	0.003	0.002	0.002	0.000	0.001	0.001	0.001	**0.110**	**0.076**	**0.566**	**0.076**
RW-1500	0.000	0.000	0.000	0.030	0.000	0.000	0.000	**0.093**	**0.154**	**0.426**	**0.118**	**0.462**
CF-500	**0.262**	**0.275**	0.028	**0.268**	**0.531**	**0.548**	0.002	**0.539**	0.000	0.000	**0.492**	0.000
CF-1000	**0.262**	**0.275**	0.028	**0.268**	**0.531**	**0.548**	**0.089**	**0.539**	0.000	0.000	**0.508**	0.000
CF-1500	**0.262**	**0.275**	0.028	**0.472**	**0.531**	**0.548**	**0.089**	0.000	0.000	0.000	**0.472**	0.000
GARCH-N	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.082**	**0.078**	**0.586**	**0.178**
GARCH-SKT	0.029	**0.275**	0.028	**0.268**	0.002	**0.548**	0.002	**0.539**	**0.506**	0.000	**0.486**	0.000
GARCH-EDF	**0.262**	0.031	0.028	**0.268**	**0.531**	0.002	0.002	**0.539**	0.000	**0.508**	**0.472**	0.000
GJR-GARCH-SKT	**0.262**	**0.275**	**0.479**	**0.268**	**0.531**	**0.548**	0.000	**0.539**	0.000	0.000	0.000	0.000
GARCH-FZ	0.002	**0.275**	0.028	0.000	0.001	**0.548**	0.002	0.001	**0.418**	0.000	**0.484**	**0.098**
GAS-1F	0.000	0.000	0.028	0.028	0.000	0.000	**0.089**	**0.089**	**0.140**	0.032	**0.504**	**0.470**
Hybrid	0.000	0.003	0.028	**0.268**	0.000	0.001	**0.089**	**0.539**	**0.062**	**0.086**	**0.504**	0.000
GAS-2F	0.000	0.003	0.000	0.000	0.000	0.001	0.000	0.000	**0.132**	**0.592**	**0.778**	**0.256**
A-GAS-1F	0.029	0.031	0.002	**0.472**	**0.090**	0.002	0.009	0.000	**0.570**	**0.492**	**0.396**	0.000
A-Hybrid	0.029	0.031	0.000	**0.268**	**0.090**	0.002	0.001	**0.539**	**0.496**	**0.492**	**0.170**	0.000

Note: This table presents the p-values of two VaR backtests and one ES backtest for four oil futures with the model estimation based on rolling windows, over the OOS period (Panel A) and the COVID-19 period (Panel B) for 16 models for $\alpha_1 = 0.1\%$. Columns 2-5 and 6-9 present the results for the Unconditional Coverage (UC) and Conditional Coverage (CC) tests for VaR. The last 4 Columns present the results of the Bootstrapping (BS) test for ES. Values greater than 0.05 (indicating no evidence against the null at 5% significance level) are in bold.

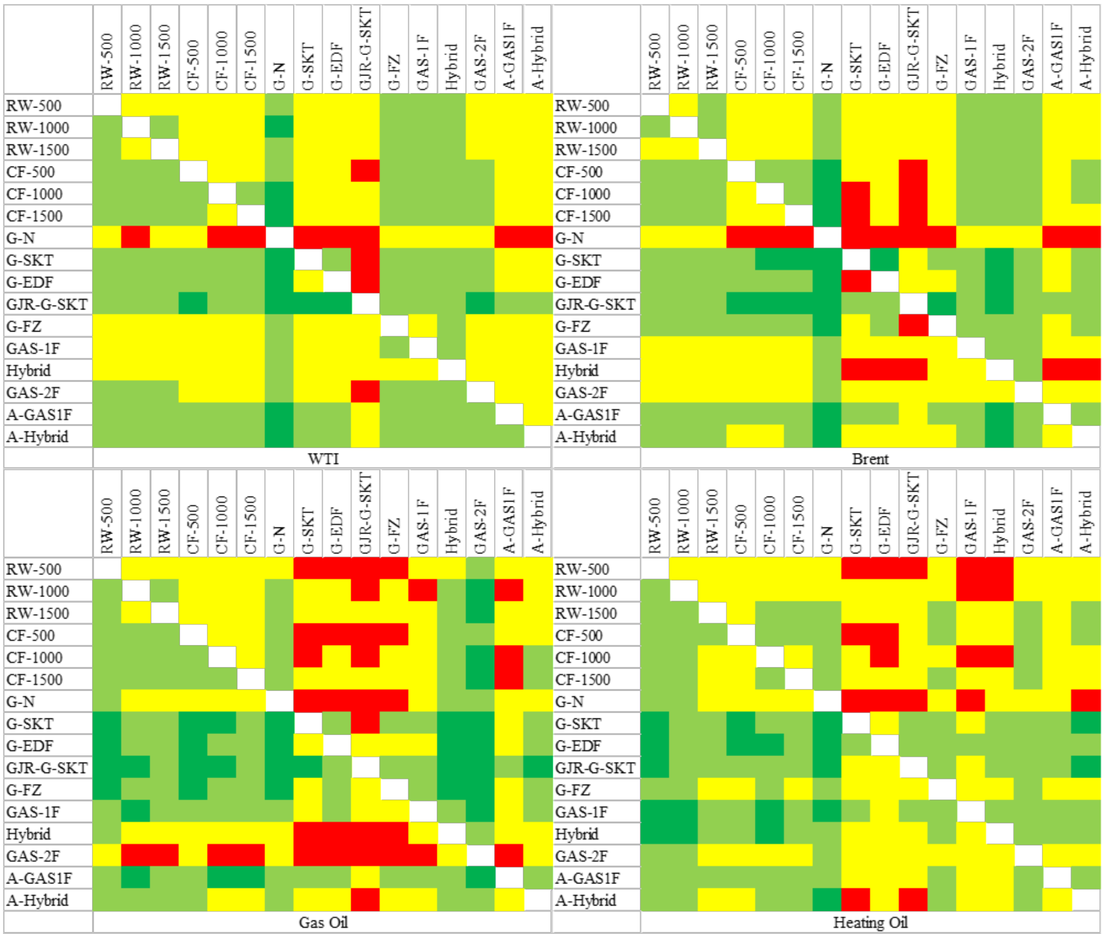
the GJR-GARCH model with skewed t innovation is the best-performing model across WTI, Brent, and GO, followed by the A-GAS models. Correspondingly, the GJR-GARCH model with skewed t innovation has the highest rank among the models, whilst the A-GAS-1F and the GARCH model with skewed t innovation are the second best among the 16 models. In terms of the COVID-19 period, the A-GAS models also outperform their competitors for all the oil futures, except GO where the A-GAS-1F is the model with the second lowest loss. The A-GAS-1F model is ranked best during the COVID-19 period while the A-Hybrid model is ranked third. According to the DM test (Figure 2.A.1), the A-GAS-1F model outperforms most of the benchmarks except the GJR-GARCH model with skewed t innovation for WTI, Brent, and GO, during the COVID-19 crisis period.

Table 2.A.2: Out-of-sample loss values and loss rankings based on a rolling window estimation

	Average loss				Loss rankings				
Panel A: Jan 2014 to Jan 2021									
	WTI	Brent	GO	HO	WTI	Brent	GO	HO	Average
RW-500	4.315	4.056	3.869	5.031	12	11	15	16	13.5
RW-1000	3.908	3.803	3.197	4.569	10	10	11	15	11.5
RW-1500	3.911	4.116	3.314	3.514	11	12	12	8	10.75
CF-500	3.599	2.994	2.998	3.474	8	6	10	7	7.75
CF-1000	3.272	3.035	2.781	3.870	6	7	8	12	8.25
CF-1500	3.310	3.175	2.730	3.575	7	9	7	10	8.25
GARCH-N	4.999	4.759	3.354	4.545	16	16	13	14	14.75
GARCH-SKT	3.087	2.660	2.394	3.026	4	3	3	3	*3.25*
GARCH-EDF	3.187	2.992	2.715	**2.833**	5	5	6	1	4.25
GJR-GARCH-SKT	**2.842**	**2.591**	**2.248**	3.037	1	1	1	4	**1.75**
GARCH-FZ	4.758	2.944	2.506	3.586	14	4	4	11	8.25
GAS-1F	4.518	4.305	2.707	*2.955*	13	13	5	2	8.25
Hybrid	4.845	4.437	3.774	3.413	15	14	14	5	12
GAS-2F	3.740	4.743	5.376	3.957	9	15	16	13	13.25
A-GAS-1F	3.047	*2.651*	*2.289*	3.466	3	2	2	6	*3.25*
A-Hybrid	*2.997*	3.123	2.841	3.548	2	8	9	9	7
Panel B: Jan 2020 to Jan 2021									
	Average loss				Loss rankings				
	WTI	Brent	GO	HO	WTI	Brent	GO	HO	Average
RW-500	10.824	11.153	9.277	9.962	10	10	15	14	12.25
RW-1000	12.517	12.562	7.458	11.196	12	13	13	16	13.5
RW-1500	13.743	14.715	8.654	5.689	13	14	14	8	12.25
CF-500	6.232	5.715	4.693	5.188	6	6	9	7	7
CF-1000	7.930	7.155	4.315	6.547	8	8	7	10	8.25
CF-1500	7.678	7.593	4.437	3.339	7	9	8	5	7.25
GARCH-N	11.964	12.135	6.474	7.238	11	12	12	13	12
GARCH-SKT	5.556	5.196	2.779	3.408	5	4	3	6	4.5
GARCH-EDF	5.484	6.345	3.918	3.151	4	7	6	3	5
GJR-GARCH-SKT	*5.287*	*4.781*	**2.555**	3.334	2	2	1	4	*2.25*
GARCH-FZ	15.815	5.385	2.798	6.148	14	5	4	9	8
GAS-1F	17.279	15.993	4.895	7.193	15	15	10	11	12.75
Hybrid	19.694	11.914	6.172	7.225	16	11	11	12	12.5
GAS-2F	10.196	18.913	15.438	10.096	9	16	16	15	14
A-GAS-1F	5.433	**4.455**	*2.768*	**2.600**	3	1	2	1	**1.75**
A-Hybrid	**4.412**	4.871	3.424	*3.096*	1	3	5	2	2.75

Note: This table presents the average losses and rankings based on average FZ0 losses with parameters obtained via rolling window estimation, for VaR and ES forecasts of four oil futures, over the OOS period (Panel A) and COVID-19 period (Panel B) for 16 models estimated for $\alpha_1 = 0.1\%$. Columns 2-5 present the average FZ0 losses, with the lowest (second lowest) in each column shown in bold (italics). Columns 6-9 present the loss rankings. The last column presents the average rank across the four series, with the best (second best) model shown in bold (italics).

Figure 2.A.1: Diebold-Mariano test based on a rolling window estimation



Note: This figure presents the Diebold–Mariano (DM) test results comparing the FZ0 losses over the OOS period from January 2014 to January 2021, for 16 models across four oil futures for $\alpha_1 = 0.1\%$. Dark green (red) blocks mean that the row model has significantly lower (greater) average loss than the column model at 10% significance level; light green (yellow) blocks mean that the row model has lower (greater) average loss than the column model, but the difference is not significant.

Notes

[1]An auxiliary level may provide information to a tail exceed a lower α, and we are investigating which level will provide better forecast on risk estimates with this information. Therefore, this is part of our research question, rather than an assumption.

[2]Fissler and Ziegel (2016) show that VaR and ES are jointly elicitable, while ES is not elicitable by itself (Gneiting, 2011).

[3]An additional auxiliary level α_3 could also be considered at higher computational cost. However, in our preliminary analysis, this does not provide considerable improvements.

[4]For the same reason, we choose to let κ_t only depend on risk measures at auxiliary level α_2.

[5]In order to obtain the risk estimates based on the specification of A-GAS models, it is reasonable to have the total loss as a function of the two loss functions.

[6]If the set of auxiliary level contains less than two candidates, then the TSCV procedure is not needed.

[7]All parameter values are obtained via fitting the model to the de-meaned returns on the S&P 500 from January 2000 to December 2020.

[8]Other values of α_2 can also be considered. Due to the computational cost, we restrict α_2 to these eight choices.

[9]The calculation of the FZ0 loss values and loss reductions are based on the Eq.(2.2.4). Also, under the Monte-Carlo simulation, the optimal α_2 is different in each replication.

[10]More backtest results are reported in the online Supplemental Appendix.

[11]According to the positive loss reduction at $\alpha_1 = 2.5\%$, our proposed models can be well estimated for less extreme values of α_1 of up to 2.5%

[12]Commodity markets are considered to be highly volatile (Del Brio et al., 2020). The proposed risk models can be applied in other markets, such as equity markets. However, due to the highly volatile nature of commodity returns, the issue of estimating risk at extreme levels is most imperative in these markets, which motivated our empirical investigation.

[13]To ensure the continuity in data, we remove the days with negative prices, market-specific non-trading days and zero returns from each return series.

[14]We use $M_1 = 30$ days to reduce computational time.

[15]The outperformance of the proposed models on both in-sample estimation and OOS forecasting implies that our model are not affected by the over-fitting issue.

[16]The online Supplementary Appendix presents the results for the robustness check.

Chapter 3

Measuring Climate Risk for equities

3.1 Introduction

As one of the most critical global challenges on this planet, climate change potentially impacts every individual, with health and social implications, but also affecting the economy and the financial system. Fossil fuels are a crucial input to production, and economic growth increases greenhouse gas emissions. The climate change attributes to those emissions and the literature shows that climate change has become a prominent risk that will potentially create substantial costs to the economy (Burke et al., 2015; Dietz et al., 2016; Lesk et al., 2016). Nonetheless, if the economic effects of climate change are as large as some studies have suggested, then, given that financial assets are ultimately supported by economic activities, the impact of climate change on financial assets could also be substantial. When the climate risk manifests itself, it is either reflected on the physical risk or transition risk. Physical risks refer to the mainly negative impact of climate and weather-related events on business operations, society, and supply chains (Tankov and Tantet, 2019). There are two sub-categories within the physical risk: acute risk and chronic risk. Extreme weather events including extreme drought and precipitation, floods, hurricanes, heatwaves, and wildfires are defined as acute risks. Chronic risks are generally considered to include: rising sea levels, rising average temperatures, and ocean acidification. In terms of the transition risk, it refers to the risk associated with a path to a low carbon economy and all related implications of fossil fuels and dependent sectors (Curtin et al., 2019).

Research on the interaction between climate change and financial economics is termed climate finance (Giglio et al., 2021). Our study contributes to the climate finance literature that investigates the impact of climate change risk on financial markets and firms. We introduce new measures of climate risk, specifically *climate Value-at-Risk* and *climate Expected Shortfall* which capture the risk in equities that stems from climate risk factors proxied by environmental scores.[1] Also, we compare the average ratios of climate Value-at-Risk and climate Expected Shortfall to total risk in several equity sectors, and we identify the sectors in which climate risk factors contribute most to total risk.

Climate change risk is a growing concern for the financial sector, and it is affecting the prices of various assets, including stocks, bonds, real estate, and more (see Bernstein et al., 2019; Goldsmith-Pinkham et al., 2019; Hong et al., 2019; Baldauf et al., 2020; Painter, 2020; Bolton and Kacperczyk, 2021; Giglio et al., 2021). Also, it is a long-term risk that poses significant challenges to investors, as it is often not effectively priced in financial markets (Andersson et al., 2016; Bansal et al., 2016). To mitigate this risk, investors need to consider the potential impact of climate change on the returns of assets. The implementation of carbon pricing can play an important role in reducing CO_2 emissions (Best et al., 2020) so that the climate risks can be alleviated, but it is also important to consider other factors, such as climate regulatory risk exposure (Seltzer et al., 2022), and the effects of weather conditions with abnormal temperature (Anttila-Hughes, 2016; Kumar et al., 2019; Choi et al., 2020). On the one hand, companies with high carbon emissions are more likely to be exposed to climate change risk, and their stock prices may be more likely to be affected by climate-related factors (Bolton and Kacperczyk, 2021). On the other hand, companies with higher environmental scores on ESG scores are likely to perform better when climate-related events occur (Engle et al., 2020; Huynh and Xia, 2021). Furthermore, climate policy uncertainty is reflected in stock option markets and can influence the social cost of carbon, as well as affecting the stock prices of firms with high exposure to climate policy (Barnett, 2017; Barnett et al., 2020; Ilhan et al., 2021). These recent literature highlight the influence of climate risk on different aspect of the financial market. In order to further contribute to this research area, Our research aims to delve deeper into the intersection of climate risk and market risk, specifically focusing on Value-at-Risk (VaR) and Expected Shortfall (ES).

Risk measures such VaR and ES have been widely used in academia and practice. VaR is one of the most popular tail risk measures that is employed to assess and manage financial risk. VaR is an estimate of the quantile of the distribution of profit and losses, and it can be measured at different levels. Due to its conceptual simplicity, VaR has become a popular risk measure of market risk and is frequently investigated (see Duffie and Pan, 1997; Dowd, 1998; Jorion, 2000; Dempster, 2002; Allen, 2012). However, since VaR ignores the shape and structure of the tail of the returns' distribution and is not a coherent risk measure (i.e. it is not subadditive), ES, as an alternative, has been proposed (Artzner, 1997; Artzner et al., 1999). It measures the expected value of the observations provided that they exceed VaR and is a coherent risk measure (Roccioletti, 2015). Due to its favourable properties, ES has consistently increased in popularity (see e.g. Chen et al., 2012; Patton et al., 2019; Taylor, 2019; Gerlach and Wang, 2020). However, the measurement of ES is inherently dependent on the value of the VaR estimate. As such, ES is not elicitable by itself, and only the (VaR, ES) tuple is elicitable (Ziegel, 2016). There is no doubt that in the recent years climate risk has become one of the most important components of total risk. One important question that arises is to what extent climate-related risks contribute to the total risk, and this is the central research question we address here.

This paper makes three main contributions. First, we investigate the relationship between stock returns and transition climate risk factors, in different return quantiles, by using firm-level environmental scores constructed by the ESG ("Environmental, Social, and Governance") data provided by Thomson Reuters ASSET4 to proxy the firms' climate risk exposure. In this Chapter, we are interested in the three components of the environmental score, because results from sub-scores provide a more intuitive and direct perspective than the overall environmental score on how to reduce the climate transition risk. We find a significant negative relationship between them in the lower quantiles of stock returns, implying that companies that face financial difficulties are affected negatively by the costs of improvements made to their environmental scores. Second, we propose novel measures (climate VaR and climate ES) that capture the market risk attributed to transition climate risk factors proxied by environmental scores. Third, we show how various sectors respond to climate risk. Our results indicate the diversity in the sensitivity of different sectors to climate risk variables. Companies in the Energy sector gain the most from improving environmental scores, whereas companies in the Health Care sector are the least cost-effective in decreasing their climate risk. The results are robust to changes to the model used to capture risk and to the levels of risk significance.

The rest of the paper is organized as follows. Section 3.2 discusses the methodology to estimate the climate risk measures. Section 3.3 introduces the firm-level data used in the empirical analysis. Section 3.4 presents the estimation results from panel data regressions. Section 3.5 reports the results of several robustness checks. Section 3.6 concludes. The online Supplemental Appendix contains additional results.

3.2 Methodology

3.2.1 Risk measures

The downside risk is captured by the left tail of stock returns' distribution. Two prevalent measures are employed to identify such risk. The first measure, VaR, is an estimate of the quantile of the distribution of profit and losses and it can be measured at different levels. Due to its conceptual simplicity, VaR has become a popular risk measure of market risk. However, VaR ignores the shape and structure of the tail of the returns' distribution and is not a coherent risk measure (i.e. it is not subadditive) (Artzner et al., 1999). Thus, a second risk measure has been introduced, ES, which measures the expected value of the observations provided that they exceed VaR; this is a coherent risk measure (Roccioletti, 2015).

VaR provides banks and financial institutions with an estimate of the minimum loss level that occurs in the worst outcomes at a given level $\alpha \in (0,1)$. Let $F_Y(\cdot \mid \Omega_{t-1})$ denote the cumulative distribution function of asset return Y_t over a time horizon (such as one day or one week) conditional on the information set Ω_{t-1}. The VaR at level α can be written directly in terms of the inverse cumulative distribution function (Duffie and Pan, 1997):

$$VaR_t^{\alpha} = F_Y^{-1}(\alpha \mid \Omega_{t-1}), \tag{3.2.1}$$

where VaR_t^{α} denotes the α-quantile of the underlying return distribution at time t. As such, Following Ziegel (2016), Nolde and Ziegel (2017), and Chen (2018), the VaR at level α at time t can be defined as:

$$VaR_t^{\alpha} = \inf\{Y_t \mid F_Y(Y_t \mid \Omega_{t-1}) \geq \alpha\}. \tag{3.2.2}$$

ES measures the expectation of return conditional on its value being less than VaR. As a coherent risk measure and due to its superior properties, ES has become increasingly popular in the risk management of banks and financial institutions. Recently, the Basel Committee on Banking Supervision (2013) proposed a transition from VaR at 1% level to ES at 2.5% level motivated by the global financial crisis in 2008. ES at level α at time t can be formally defined as (see Acerbi and Tasche, 2002):

$$ES_t^{\alpha} = \mathbb{E}[Y_t \mid Y_t \leq VaR_t^{\alpha}, \Omega_{t-1}]. \tag{3.2.3}$$

Since the generalized autoregressive conditional heteroskedastic (GARCH) model of Bollerslev (1986) and its variants (Nelson, 1991) capture the time-varying volatility feature, they are widely used to forecast VaR and ES in the literature. We also employ the GARCH model with skewed t distribution of Hansen (1994) for our estimation of risk measures. The model is specified as follows:

$$\begin{aligned} v_t &= \mu_t + a\ \sigma_t, \quad \text{where } a = F_\eta^{-1}(\alpha), \\ e_t &= \mu_t + b\ \sigma_t, \quad \text{where } b = \mathbb{E}\ [\eta_t | \eta_t \leq a], \\ Y_t &= \sigma_t\ \eta_t, \quad \eta_t \sim iid\ F_\eta(0,1), \\ \sigma_t^2 &= \omega + \delta\sigma_{t-1}^2 + \gamma Y_{t-1}^2, \end{aligned} \tag{3.2.4}$$

where σ_t^2 is the conditional variance which follows a GARCH(1, 1) process, η_t is the standardized residual that follows the skewed t distribution $F_\eta(0,1)$ and Y_t is the de-meaned daily returns. We transform the daily VaR and ES to monthly estimates by multiplying average daily risk measures in the given month by the square root of 21. There are many other ways to estimate VaR and ES. We provide the robustness check using alternative estimation of VaR and ES in Section 3.5.

3.2.2 Climate VaR and ES

We employ the environmental component (denoted as E-score) of the ESG score in our study, given that it is related to the environmental factors and captures the effects of climate-related issues on companies. The E-score is comprised of three sub-scores: the *Emission* score, *Innovation* score, and *Resource Use* score. Specifically, the Emission score reflects the extent to which a firm is committed to reducing environmental emissions in its production and operational processes; the Innovation score measures a firm's capacity to create new market opportunities through environmental technologies and processes, or eco-designed products; the Resource Use score reflects a firm's performance and capacity to reduce the amount of natural resources it uses and improve its supply chain management. Taken together, these sub-components provide a comprehensive view of a firm's environmental performance and can help investors make informed decisions about the long-term sustainability and financial performance of a company. Thus, instead of directly revealing the link between this environmental pillar and the downside risks, we consider these three sub-components of the E-score in order to quantify the downside risks attributed to the climate risk factors.

To determine the extent to which the risk presented by climate factors affects the VaR and ES of the equity returns, we begin our analysis by investigating the link between risk measures and environmental scores in various sectors. For every sector, we estimate the following panel data regression:

$$\begin{aligned} Downside\ Risk_{i,t} =& \beta_0 + \beta_1 Emission_{i,t} + \beta_2 Innovation_{i,t} + \\ & \beta_3 Resource_{i,t} + \beta_4 Controls_{i,t-1} + \delta_i + \gamma_t + \epsilon_{i,t}, \end{aligned} \tag{3.2.5}$$

where the $Downside\ Risk_{i,t}$ represents one of the two risk measures ($VaR_{i,t}$ and $ES_{i,t}$) of the firm i in month t at 1% level; $Emission_{i,t}$, $Innovation_{i,t}$ and $Resource_{i,t}$ measure the Emission, Innovation and Resource Use scores, respectively, of firm i in month t; $Controls_{i,t-1}$ is a vector of control variables that may affect downside risk, including size, M/B, leverage, ROE, and investment.[2] We include firm fixed effect (δ_i) and year-month fixed effect (γ_t). We obtain $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, and these capture the effects of the environmental risk factors on VaR and ES. Also, we report the heteroskedasticity-consistent standard errors of White (1980).

In the following, we provide the definition for Climate VaR and ES, which are the VaR and ES of the stock returns of a firm, attributed to environmental scores. Based on Eq.(3.2.5), the Climate VaR and ES of firm i in month t are calculated as:

$$\begin{aligned} Climate\ Downside\ Risk_{i,t} = &\hat{\beta}_1 Emission_{i,t} + \hat{\beta}_2 Innovation_{i,t} + \\ &\hat{\beta}_3 Resource_{i,t}, \end{aligned} \tag{3.2.6}$$

where β parameters measure the association between market risk and environmental scores. If the β is negative (positive), an increase in the Emission score, Innovation score, or Resource Use score increases (decreases) the risk.[3] Additionally, we define the portion of VaR or ES attributable to environmental scores as follows:

$$Climate\ Risk\ Ratio_{i,t} = \frac{Climate\ Downside\ Risk_{i,t}}{Downside\ Risk_{i,t}}. \tag{3.2.7}$$

When the sign of the ratio is negative, the effort spent on the improvement of these three environmental scores reduces the riskiness of the firm. When it is positive, the cost associated with the improvement of the environmental scores leads to an increase in the firm's downside risk.

3.2.3 Quantile regression with penalized fixed effect for panel data

In the recent literature, several environmental proxies have been shown to affect stock returns (Engle et al., 2020; Bolton and Kacperczyk, 2021; Hsu et al., 2023). Here we employ the quantile regression proposed by Koenker (2004) using panel data to discover the relationship between stock returns and environmental scores at various quantiles. To determine how environmental scores influence returns at different quantiles of their distribution, we first investigate the following standard linear panel regression model:

$$y_{i,t} = x_{i,t}^{\top}\beta + \delta_i + \epsilon_{i,t} \;\; t = 1, ..., T_i, \; i = 1, ..., n, \tag{3.2.8}$$

where $y_{i,t}$ indicates the firm's stock return, $x_{i,t}$ is a vector of variables including the Emission score, Innovation score, Resource Use score, and the lagged one-month size, M/B, leverage, ROE, and investment. δ_i represents the firm fixed effect, and $\epsilon_{i,t}$ is the error term. The subscript i indexes the firm, while the subscript t indexes the time.[4] The following model is then considered for the conditional quantile functions (at quantile τ) of the returns in month t of the ith firm y_{it}:

$$Q_{y_{it}}(\tau \mid x_{it}) = x_{it}^{\top}\beta(\tau) + \delta_i, \;\; t = 1, ..., T_i, \; i = 1, ..., n. \tag{3.2.9}$$

To simultaneously estimate Eq. (3.2.9) for several quantiles, we perform the following optimization:

$$\min_{(\beta,\delta)} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{t=1}^{T_i} w_k \rho_{\tau_k}(y_{it} - x_{it}^{\top}\beta(\tau_k) - \delta_i), \tag{3.2.10}$$

where $\rho_\tau(\epsilon) = \epsilon(\tau - I(\epsilon < 0))$ denotes the piecewise linear quantile loss function of Koenker and Bassett (1978). The weights w_k control the relative impact of the q quantiles $\{\tau_1, ..., \tau_q\}$ on the estimation of the parameters.

The estimation of β and the firm fixed-effect δ_i can be improved by reducing the unconstrained $\hat{\delta}_i$'s toward a common value. To achieve that, we employ the ℓ_1 penalty, $P(\delta) = \sum_{i=1}^{n} |\delta_i|$ in addition to Eq. (3.2.10).[5] Then, we obtain the estimators by solving the penalized version of Eq. (3.2.10):

$$\min_{(\beta,\delta)} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{t=1}^{T_i} w_k \rho_{\tau_k}(y_{it} - x_{it}^{\top}\beta(\tau_k) - \delta_i) + \lambda \sum_{i=1}^{n} |\delta_i|, \;\; \lambda > 0, \tag{3.2.11}$$

where λ is the penalty term. For $\lambda \mapsto 0$ we obtain the fixed effects estimator described in Eq. (3.2.10), while as $\lambda \mapsto \infty$ the $\hat{\delta} \mapsto 0$ for all $i = 1, 2, ..., n$ and we obtain an estimate of the model with the fixed effects eliminated.

3.3 Data

In this section, we describe all datasets used in the empirical analysis. To avoid the potential structural break during the COVID-19 period, our primary database ranges from January 2003 to December 2019 and is primarily comprised of three datasets obtained from Thomson Reuters ASSET4 ESG database, Compustat, and CRSP. Thomson Reuters ASSET4 provides data on environmental scores, Compustat provides data on corporate fundamentals, and CRSP provides data on stock returns. We implement the matching using CUSIP as the main identifier, and the ultimate matching produces 802 unique firms and 58311 firm-month observations.[6]

According to Section 3.2.2, we measure firm-level environmental performance using the Emission scores, Innovation scores, and Resource Use scores under the environmental pillar of the Thomson Reuters ASSET4 ESG scores. Calculated at the firm-quarter level, our control variables are defined as follows. $Size$ is the natural logarithm of the firm's market capitalization. M/B is the firm's market capitalization divided by its book value. $Leverage$ is the book leverage of the firm defined as the sum of long-term and short-term debt divided by common equity. ROE is the firm's earning performance. $Investment$ is the natural logarithm of one plus firm's capital expenditure (to avoid the natural logarithm of zero). To mitigate the impact of outliers, M/B, $Leverage$, and ROE are winsorized at 1% level. We note that firms in various sectors have diverse responses to environmental scores. Hence, we report the summary statistics of the sample with respect to the FTSE/DJ Industry Classification Benchmark (ICB) in Table

3.3.1. Telecommunications has the lowest average return with a value of 0.633%, while Technology has the highest average return (2.080%), followed by Health Care (1.619%). Energy has the greatest Emission and Innovation scores, with respective values of 57.860 and 53.371. Health Care has the highest Resource Use score (63.459), but the lowest Innovation score (41.626). The lowest Emission and Resource Use scores are reported for Industrials, which are 38.544 and 42.610, respectively.

3.4 Results

3.4.1 Quantile regression results

Given the definition of VaR and ES, they both capture the potential loss of the asset, which is equivalent to the left-tail or low quantile of the return distribution. Prior to examining how market risk affected by climate transition risk, it is crucial to explore the link between climate transition risk and asset returns. Therefore, we begin our analysis by investigating the relationship between stock returns in different quantiles and the Emission score, the Innovation score, and the Resource Use score, by employing the quantile regression described in Section 3.2.3. Table 3.4.2 reports the panel regression results for quantiles $\tau \in \{1\%, 5\%, 10\%, 30\%, 50\%, 70\%, 90\%, 95\%, 99\%\}$, where all quantiles are assigned with equal weights when estimating using Eq.(3.2.11). For quantiles below 50%, significantly negative signs are observed for all three environmental scores, with the exception of the Emission score at 10% and 30% quantiles. The overall trend is that the effect is negative for lower quantiles and positive for higher quantiles, and is more

Table 3.3.1: Regression results for environmental scores on ES

Sector	*Return* (%)	*Emission*	*Innovation*	*Resource*	*Size*	*M/B**	*ROE**	*Leverage**	*Investment*	*NumComp*
Basic Materials	1.051	41.787	44.425	44.989	7.838	2.652	0.030	1.427	4.234	62
Consumer Discretionary	1.134	44.398	45.990	50.120	8.610	4.621	0.049	1.541	4.672	125
Consumer Staples	0.845	43.764	45.654	53.316	8.873	7.178	0.068	1.690	4.613	61
Energy	0.988	57.860	53.371	56.902	8.761	1.126	0.001	0.558	5.473	28
Financials	1.398	52.126	41.626	49.032	9.298	5.905	0.066	1.444	2.895	65
Health Care	1.619	56.961	46.748	63.459	9.624	7.770	0.020	0.740	4.523	31
Industrials	1.236	38.544	45.516	42.610	8.258	5.322	0.054	1.765	3.940	188
Real Estate	1.017	53.846	43.737	51.256	8.467	2.498	0.019	1.374	0.738	74
Technology	2.080	52.132	52.158	54.930	9.008	4.660	0.031	0.436	4.201	88
Telecommunications	0.633	48.236	46.678	53.300	8.948	2.426	0.011	1.175	4.970	22
Utilities	1.198	56.096	44.961	49.170	8.827	2.015	0.025	1.264	6.039	58

Note: This table reports averages (for monthly frequency) of the variables employed in the regressions in this study reported for 11 different sectors listed in the first column. The sample period is from January 2003 to December 2019. *Return* represents average monthly return of the sector (in percentages). *Emission*, *Innovation*, and *Resource* indicate, respectively, the Emission score, Innovation score, and Resource Use score. *Size* is the natural logarithm of market capitalization in $ million. M/B denotes the market value of equity divided by its book value. *ROE* is the return on equity. *Leverage* is the total debt (long-term and short-term) divided by the total stockholders' equity. *Investment* is the natural logarithm of the capital expenditures in $ million. *NumComp* represents the number of companies in the sector. Variables followed by * are winsorized at 1%.

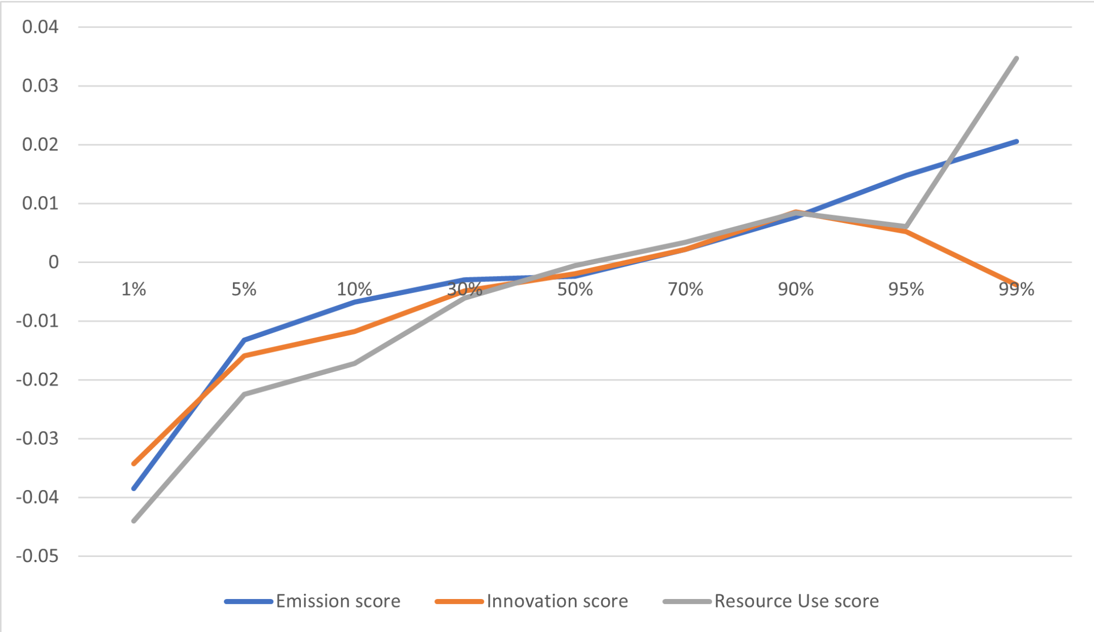
pronounced for lower quantiles. The signs of the control variables are generally consistent with the literature. Figure 3.4.1 illustrates the values of the coefficients of the three environmental scores ranging from $\tau = 1\%$ to $\tau = 99\%$. At the 1% quantile, environmental scores have the most negative effect on the stock returns. This effect eliminates when the quantile reaches the 50% quantile, at which point this effect disappears and becomes a positive effect, except for the Innovation score at the 99% quantile, which has a negative effect. When companies struggle, then the costs associated with improving their E-scores bring additional burdens and so improving the E-scores reduces overall returns. The effect is opposite when companies do well, in such instances improving the E-scores increases company returns.

Table 3.4.2: Quantile regression results for returns and environmental scores

	Quantiles								
Variable	1%	5%	10%	30%	50%	70%	90%	95%	99%
Emission	-0.039**	-0.013*	-0.007	-0.003	-0.002	0.002	0.008	0.015	0.021**
	(0.015)	(0.008)	(0.006)	(0.004)	(0.002)	(0.003)	(0.006)	(0.007)	(0.030)
Innovation	-0.034**	-0.016***	-0.012***	-0.005**	-0.002	0.002	0.009**	0.005	-0.004
	(0.015)	(0.006)	(0.004)	(0.002)	(0.001)	(0.002)	(0.004)	(0.008)	(0.028)
Resource	-0.044**	-0.0225**	-0.017**	-0.006**	-0.001	0.003***	0.008	0.006	0.035
	(0.021)	(0.010)	(0.008)	(0.003)	(0.001)	(0.001)	(0.007)	(0.012)	(0.029)
Size	4.775***	2.911***	2.172***	0.935***	0.289***	-0.372***	-1.713***	-2.660***	-5.365***
	(0.262)	(0.169)	(0.154)	(0.077)	(0.035)	(0.032)	(0.126)	(0.158)	(0.365)
M/B	0.023	0.010	0.002	0.001	0.001	0.000	-0.001	-0.002	0.010
	(0.016)	(0.011)	(0.010)	(0.005)	(0.003)	(0.003)	(0.007)	(0.011)	(0.033)
ROE	-0.728	0.013	0.002	-0.114	0.024	0.010	-0.242*	-0.437	-0.826***
	(0.504)	(0.379)	(0.257)	(0.188)	(0.147)	(0.144)	(0.136)	(0.154)	(0.123)
Leverage	-0.215	-0.115*	-0.012	-0.000	-0.001	0.003	0.044	0.128	0.260**
	(0.156)	(0.064)	(0.041)	(0.013)	(0.006)	(0.009)	(0.036)	(0.064)	(0.115)
Investment	-0.279	-0.427***	-0.363***	-0.252***	-0.183***	-0.122**	-0.013	0.062	0.059
	(0.200)	(0.128)	(0.107)	(0.049)	(0.030)	(0.050)	(0.115)	(0.137)	(0.277)

Note: This table presents the results of the panel quantile regression with penalized fixed firm effects for the panel data of returns and environmental scores for 11 sectors during the sample period from January 2003 to December 2019. The quantiles considered are $1\%, 5\%, 10\%, 30\%, 50\%, 70\%, 90\%, 95\%,$ and 99%. All control variables are lagged by one month. The standard errors are reported in parenthesis, *, **, and *** denote statistical significance at 10%, 5%, and 1% levels, respectively.

Figure 3.4.1: Effect of the environmental scores on returns at different quantiles



Note: This figure presents the effect of the the Emission score, Innovation score, and Resource Use score on returns at quantiles $\tau \in \{1\%, 5\%, 10\%, 30\%, 50\%, 70\%, 90\%, 95\%, 99\%\}$.

3.4.2 Climate VaR and ES results

The quantile regression results of Section 3.4.1 show that there is a differential effect of the environmental scores on the returns, depending on which quantile the returns falls into. This subsection examines the relationship between downside risk (VaR and ES) and environmental scores. We collect daily stock returns from January 2003 to December 2019 using CUSIP from CRSP as described in Section 4.3. Then, the firm-month VaR and ES at 1% level are estimated using the specification in Section 3.2.1. We present the average monthly VaR and ES across several sectors in columns 1 and 2 of Table 3.4.3. Real Estate and Utilities are the sectors with the lowest average VaR and ES, whereas Energy is the sector with the highest total risk.

Table 3.4.3: Summary statistics for VaR and ES estimates at 1% level

Sector	VaR	ES	Climate VaR	Climate ES
Basic Materials	-27.993	-37.965	0.401	0.606
Consumer Discretionary	-28.306	-39.501	-0.258	-0.784
Consumer Staples	-22.031	-31.454	2.217	3.186
Energy	-30.231	-40.293	6.642	9.346
Financials	-21.000	-27.844	0.372	0.201
Health Care	-21.370	-30.283	-4.928	-7.134
Industrials	-25.150	-34.763	-0.583	-0.684
Real Estate	-17.071	-22.260	-0.329	-0.430
Technology	-27.349	-38.528	2.035	2.798
Telecommunications	-28.696	-41.374	-1.656	-2.368
Utilities	-16.920	-22.357	1.210	1.681

Note: This table reports the average firm-month total VaR and ES as well as climate VaR and ES (in percentages) for 11 sectors during the period from January 2003 to December 2019. In columns 1 and 2, average VaR and ES estimates at 1% level are presented. Average climate VaR and ES calculated using Eq.(3.2.6) are reported in columns 3 and 4. The negative coefficients of environmental scores in Table 3.4.4 and 3.4.5 may lead to positive Climate VaR or ES estimates. A positive (negative) Climate VaR or ES means that the environmental scores contribute to a reduction (increase) in the total risk.

To reveal the effects of environmental scores on downside risk, we regress the VaR and ES at 1% level on the Emission score, Innovation score, Resource Use score, along with firm-level control variables. The results are presented in Table 3.4.4 and Table 3.4.5 for VaR and ES, respectively.[7] The Energy and Utilities sectors have only positive coefficients across all scores, indicating that an improvement in any one of these environmental scores of firms in these two sectors leads to a reduction in the total risk of the firms. Health Care, however, has solely negative coefficients on the environmental scores, which indicates that as the environmental scores increase, the firms' total risks increase proportionally. In other words, the companies' investments in improving their environmental scores reduce their total risk in the Energy and Utilities sectors, whilst it increases their total risk in the Health Care sector. Other sectors have coefficients with mixed signs associated with the three environmental scores. Due to the differences of sectors, some sectors benefit from increases in the individual scores but are negatively affected by others. For instance, firms in the Industrials sector have their risk affected positively by their Emission score but negatively by their Innovation score and Resource Use score.

Figure 3.4.2 displays the heatmaps of the statistical significance and economic significance of VaR with respect to the three environmental scores. According to the value of the coefficients, sectors including Consumer Staples, Energy, and Utilities benefit from the improvement in all of the three environmental scores. The Innovation score has a positive and statistically significant effect on the total risk of the companies in these three sectors. This effect is also observed for Resource Use Score in the Consumer Staples and Energy sectors. However, the

Table 3.4.4: Regression results for environmental scores on VaR

Variable	(1) Basic Materials	(2) Consumer Discretionary	(3) Consumer Staples	(4) Energy	(5) Financials	(6) Health Care	(7) Industrials	(8) Real Estate	(9) Technology	(10) Telecommunications	(11) Utilities
Emission	-0.027***	0.014**	0.002	0.014	0.009	-0.053***	0.015***	-0.016***	0.044***	0.050***	0.006
	(0.007)	(0.006)	(0.005)	(0.014)	(0.014)	(0.010)	(0.004)	(0.006)	(0.007)	(0.011)	(0.006)
Innovation	-0.004	-0.001	0.014***	0.031***	-0.030***	-0.006	-0.017***	0.020***	0.004	0.018**	0.016***
	(0.006)	(0.005)	(0.003)	(0.008)	(0.010)	(0.011)	(0.004)	(0.005)	(0.006)	(0.008)	(0.004)
Resource	0.038***	-0.017***	0.028***	0.074***	0.023**	-0.026***	-0.010**	-0.007	-0.008	-0.092***	0.003
	(0.007)	(0.006)	(0.005)	(0.013)	(0.011)	(0.010)	(0.004)	(0.005)	(0.008)	(0.018)	(0.005)
Size	4.288***	5.801***	4.236***	6.510***	7.632***	6.602***	5.479***	3.670***	2.477***	1.030**	8.209***
	(0.657)	(0.290)	(0.382)	(0.575)	(0.808)	(0.704)	(0.353)	(0.582)	(0.269)	(0.448)	(1.915)
M/B	0.527***	-0.022**	0.003	-0.016	-0.007	0.053***	-0.005*	1.681***	-0.039***	0.531***	-0.352
	(0.080)	(0.011)	(0.003)	(0.081)	(0.017)	(0.012)	(0.003)	(0.196)	(0.012)	(0.112)	(0.223)
ROE	1.869**	0.869*	0.211***	1.429	4.752***	4.809***	0.103	0.481	-0.521	4.957***	4.229**
	(0.861)	(0.524)	(0.071)	(1.281)	(1.646)	(1.287)	(0.203)	(3.705)	(0.321)	(1.679)	(1.970)
Leverage	-0.517***	0.014	-0.019	0.047	-0.575***	-0.910***	0.003	-3.456***	0.102**	-1.494***	0.434
	(0.073)	(0.031)	(0.020)	(0.091)	(0.106)	(0.142)	(0.003)	(0.378)	(0.045)	(0.254)	(0.430)
Investment	-0.511***	0.061	-0.115	0.775**	-0.251**	-0.742***	0.074	0.571**	-0.160	-1.754***	-0.519**
	(0.189)	(0.112)	(0.097)	(0.344)	(0.098)	(0.183)	(0.094)	(0.272)	(0.098)	(0.338)	(0.259)
Constant	-59.730***	-79.160***	-60.890***	-104.100***	-98.490***	-76.470***	-71.030***	-49.790***	-51.450***	-23.140***	-89.160***
	(5.229)	(2.614)	(3.639)	(5.858)	(8.051)	(6.659)	(3.040)	(5.204)	(2.289)	(4.267)	(17.830)
Observations	4,009	8,590	5,092	1,789	4,699	2,125	13,280	3,680	7,574	1,742	5,489
Adjusted R-squared	0.766	0.738	0.792	0.871	0.866	0.800	0.761	0.870	0.711	0.793	0.619
Year-Month F.E.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm F.E.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Note: This table provides panel regression estimates for the impact of environmental scores on VaR. Regressions are estimated at sector level. The independent variables are *Emission*,*Innovation*, and *Resource*. All control variables are lagged by one month. The sample duration extends from January 2003 to December 2019. All regressions include the year-month fixed effect and firm fixed effect. Huber-White robust standard errors are used in the regression and reported in parentheses. ** and *** denote statistical significance at 5% and 1% levels, respectively.

Table 3.4.5: Regression results for environmental scores on ES

Variable	(1) Basic Materials	(2) Consumer Discretionary	(3) Consumer Staples	(4) Energy	(5) Financials	(6) Health Care	(7) Industrials	(8) Real Estate	(9) Technology	(10) Telecommunications	(11) Utilities
Emission	-0.037***	0.019**	0.004	0.025	0.006	-0.072***	0.021***	-0.022***	0.064***	0.072***	0.009
	(0.010)	(0.008)	(0.008)	(0.019)	(0.020)	(0.014)	(0.006)	(0.007)	(0.010)	(0.016)	(0.008)
Innovation	-0.004	-0.007	0.020***	0.036***	-0.042***	-0.010	-0.021***	0.026***	0.007	0.026**	0.022***
	(0.008)	(0.007)	(0.004)	(0.010)	(0.015)	(0.016)	(0.005)	(0.006)	(0.00844)	(0.012)	(0.006)
Resource	0.052***	-0.026***	0.039***	0.105***	0.033**	-0.040***	-0.013**	-0.008	-0.016	-0.133***	0.004
	(0.010)	(0.008)	(0.008)	(0.018)	(0.015)	(0.014)	(0.006)	(0.007)	(0.011)	(0.026)	(0.006)
Size	5.879***	8.120***	6.145***	8.632***	9.975***	9.500***	7.557***	4.942***	3.538***	1.629**	11.000***
	(0.908)	(0.409)	(0.555)	(0.789)	(1.104)	(1.010)	(0.500)	(0.773)	(0.394)	(0.645)	(2.560)
M/B	0.690***	-0.031**	0.005	-0.026	-0.013	0.076***	-0.006	2.199***	-0.056***	0.780***	-0.458
	(0.107)	(0.015)	(0.005)	(0.111)	(0.027)	(0.018)	(0.004)	(0.257)	(0.017)	(0.161)	(0.298)
ROE	2.338**	1.230*	0.304***	1.837	6.250***	6.865***	0.148	1.137	-0.750	7.183***	5.675**
	(1.129)	(0.728)	(0.0994)	(1.732)	(2.196)	(1.846)	(0.279)	(4.693)	(0.476)	(2.400)	(2.631)
Leverage	-0.675***	0.018	-0.030	0.069	-0.702***	-1.295***	0.004	-4.562***	0.150**	-2.200***	0.564
	(0.097)	(0.042)	(0.028)	(0.123)	(0.137)	(0.204)	(0.005)	(0.495)	(0.066)	(0.368)	(0.576)
Investment	-0.686***	0.020	-0.162	0.916**	-0.306**	-1.042***	0.071	0.738**	-0.266*	-2.474***	-0.757**
	(0.253)	(0.153)	(0.140)	(0.465)	(0.130)	(0.256)	(0.125)	(0.354)	(0.141)	(0.478)	(0.351)
Constant	-81.600***	-109.700***	-87.650***	-137.700***	-128.900***	-109.100***	-97.620***	-66.050***	-72.420***	-34.640***	-118.60***
	(7.205)	(3.663)	(5.280)	(8.005)	(10.99)	(9.527)	(4.307)	(6.914)	(3.321)	(6.094)	(23.840)
Observations	4,009	8,590	5,092	1,789	4,699	2,125	13,280	3,680	7,574	1,742	5,489
Adjusted R-squared	0.770	0.767	0.812	0.882	0.863	0.816	0.770	0.872	0.716	0.813	0.631
Year-Month F.E.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm F.E.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

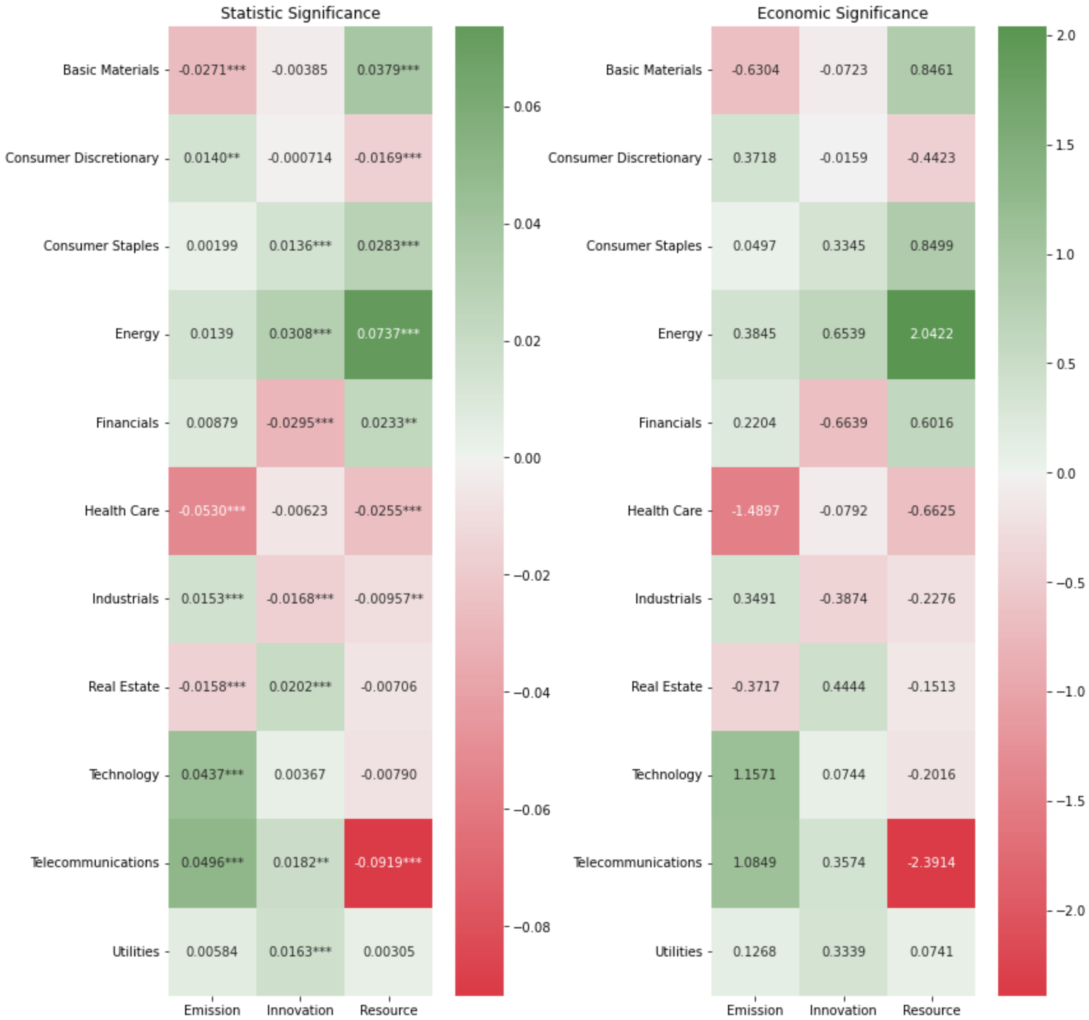
Note: This table provides estimates of the effect of environmental scores on ES based on panel regressions. Regressions are estimated at sector level. The independent variables are *Emission*, *Innovation*, and *Resource*. All control variables are lagged by one month. The sample period is from January 2003 to December 2019. All regressions include the year-month fixed effect and firm fixed effect. Huber-White robust standard errors are used in the regression and reported in parentheses. ** and *** denote statistical significance at 5% and 1% levels, respectively.

negative signs of the coefficients of the three environmental scores in the Health Care sector indicate that the additional expenditures made by companies to improve their environmental scores raise their total risk. Economical significance can also be observed: one-standard-deviation increase in the Resource Use score of companies in the Energy sector leads to a reduction of 2.042% in their total risk, whereas one-standard-deviation increase in the Resource Use score of companies in the Telecommunication sector is associated with a 2.392% increase in their total risk.

Climate VaR and ES are computed based on Eq. (3.2.6), and the results are presented in columns 3 and 4 in Table 3.4.3. In the Energy sector, the average Climate VaR (ES) is the most positive at 6.642% (9.346%), which implies that the environmental scores lead to a reduction of total VaR (ES). On the contrary, the VaR and ES of firms in Health Care attributed to environmental scores are the highest in absolute value. The cost associated with improving the environmental scores leads to an increase in the firms' downside risk in this sector. A similar effect can be seen in the Telecommunication sector.

We employ the climate risk measure proposed in Eq. (3.2.7) to demonstrate the extent to which the environmental scores affect the total downside risk of the firms. The summary statistics of the climate risk ratio for VaR and ES for different sectors are reported in Table 3.4.6. A negative (positive) sign in the mean value of the climate risk ratio indicates that, on average, improvements in the environmental scores reduce (increase) the total risk of the firm. Sectors including Basic Materials, Consumer Staples, Energy, Financials, Technology, and Utilities benefit from the effort spent on increasing the companies' environmental

Figure 3.4.2: Heatmaps of statistical and economic significance for VaR



Note: This figure presents heatmaps of the Statistical significance (left) and Economic significance (right) of the Emission score, Innovation score, and Resource Use score for VaR from 11 sectors during the sample period from January 2003 to December 2019. The statistical significance is represented by the coefficients of environmental scores in Table 3.4.4. *, **, and *** denote statistical significance at 10%, 5%, and 1% levels, respectively. Economic significance is defined as the percentage change in total VaR associated with an increase of one standard deviation in the specified environmental score. In both heatmaps, red (green) boxes indicate that an improvement in the specified environmental score increases (decreases) risk scores, and the proportion of total VaR reduced by environmental scores ranges from 1.798% to 26.996%. Sectors such as Consumer Discretionary, Health Care, Industrials, Real Estate, and Telecommunications are negatively affected by the increases in the companies' environmental scores, but the effect on their total VaR is less than 7.2%, with the exception of Health Care, which is characterized by VaR increases of 26.499% on average, due to the companies' environmental scores. Similar results can be found for ES.

Table 3.4.6: Summary statistics of climate risk ratio for VaR and ES at 1% level

	Mean		Std		Max		Min	
Sector	VaR	ES	VaR	ES	VaR	ES	VaR	ES
Basic Materials	-1.798	-1.947	3.095	3.088	3.658	3.643	-11.050	-10.428
Consumer Discretionary	1.026	2.249	1.301	1.744	6.161	9.259	-1.092	-0.679
Consumer Staples	-12.382	-12.771	7.997	8.515	-1.240	-1.217	-32.989	-36.480
Energy	-26.996	-29.397	18.960	21.607	-4.905	-5.120	-76.727	-86.979
Financials	-1.896	-0.805	4.032	4.084	6.510	9.748	-11.419	-10.604
Health Care	26.499	27.671	15.638	16.909	69.415	75.364	1.987	2.130
Industrials	2.585	2.223	2.139	2.020	9.570	8.892	-2.476	-2.753
Real Estate	1.879	1.900	3.590	3.660	10.070	10.118	-8.939	-8.909
Technology	-8.086	-7.954	4.949	4.972	-0.410	-0.264	-23.312	-23.123
Telecommunications	7.141	7.382	8.882	9.614	29.765	33.169	-9.396	-9.684
Utilities	-7.776	-8.314	3.672	3.997	-1.459	-1.606	-15.918	-17.169

Note: This table presents the summary statistics of the climate risk ratio for VaR and ES (in percentages) for 11 sectors from January 2003 to December 2019. The mean values and standard deviations of the ratio appear in columns 1-2 and 3-4, while the maximum and minimum values of the ratio appear in columns 5-6 and 7-8.

To visually illustrate the fraction of VaR and ES that is attributable to the environmental scores, we display summary statistics of the climate risk ratio of VaR and ES in Figure 3.4.3, and sort the climate risk ratio of different sectors in descending order in both panels. In the Health Care sector, the climate risk factors contribute approximately 27% on average to the total VaR and ES, the 95% quantile of the ratio for VaR is 49.956% and for ES it is 53.454%. On the other hand, in the Energy sector, these factors can reduce VaR or ES by about

28% on average and the 5% quantile of the ratio for VaR is -65.294% and for ES it is -71.872%. The ranking of sectors including Health Care, Telecommunication, Consumer Staples, and Energy are the same in both Figure 3.4.3a and Figure 3.4.3b. As the climate risk variables have a smaller impact on companies in the other sectors, their climate risk ratio is around zero.

3.5 Robustness checks

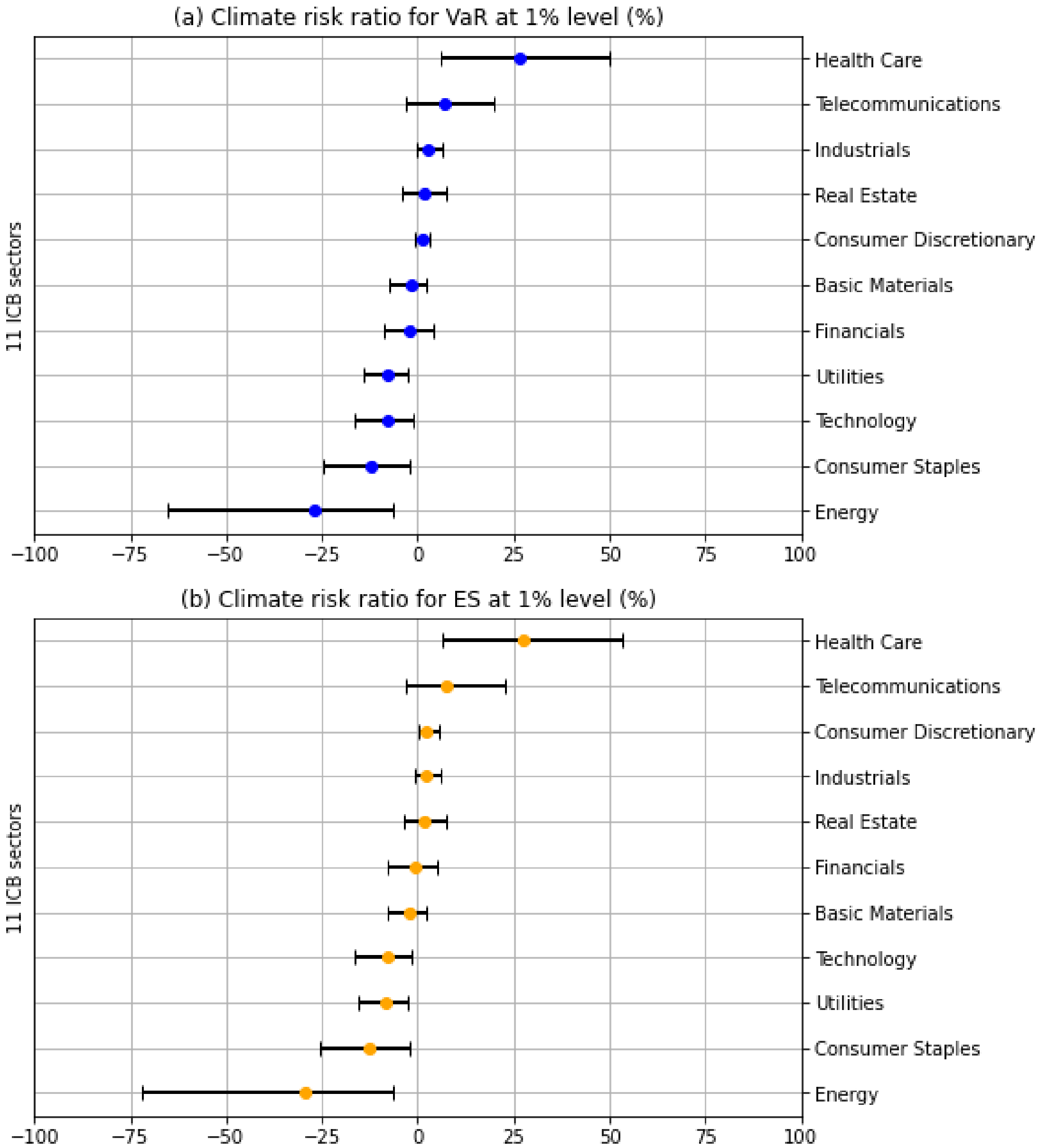
3.5.1 Alternative VaR and ES models

To investigate the result's sensitivity to the risk estimation model, We repeat our previously presented climate risk esitmation methodology using alternative VaR and ES models. Specifically, we consider the GJR-GARCH model with skewed t innovation and the GARCH-FZ model of Patton et al. (2019). Table 3.5.7 depicts the various climate risk ratios for the three previously discussed models. We notice that the various risk models yield similar but slightly different values for the climate risk ratio. When it comes to the ranking of the sectors based on their climate risk ratios, however, there is a high degree of consistency among the various models, with the ratios remaining mostly unaffected.

3.5.2 Alternative risk levels

After the 2007–2008 financial crisis, the Basel Committee on Banking Supervision (2013) proposed a transition from 1% VaR to 2.5% ES. In addition to VaR and ES at 1%, different risk levels are therefore explored in this robustness check. We employ VaR at 2.5% and 5% levels estimated from the GARCH model with

Figure 3.4.3: Climate risk ratio for 11 sectors at 1% level



Note: This figure presents the climate risk ratio (in percentages) for 11 sectors at 1% level. The ratios for Var and ES are displayed in (a) and (b), respectively. The left and right boundaries of the error bar for each sector are the 5 percent and 95 percent quantiles of the ratio, while the coloured marker represents the mean value. The sectors in the panel are ordered in descending order of the climate risk ratio

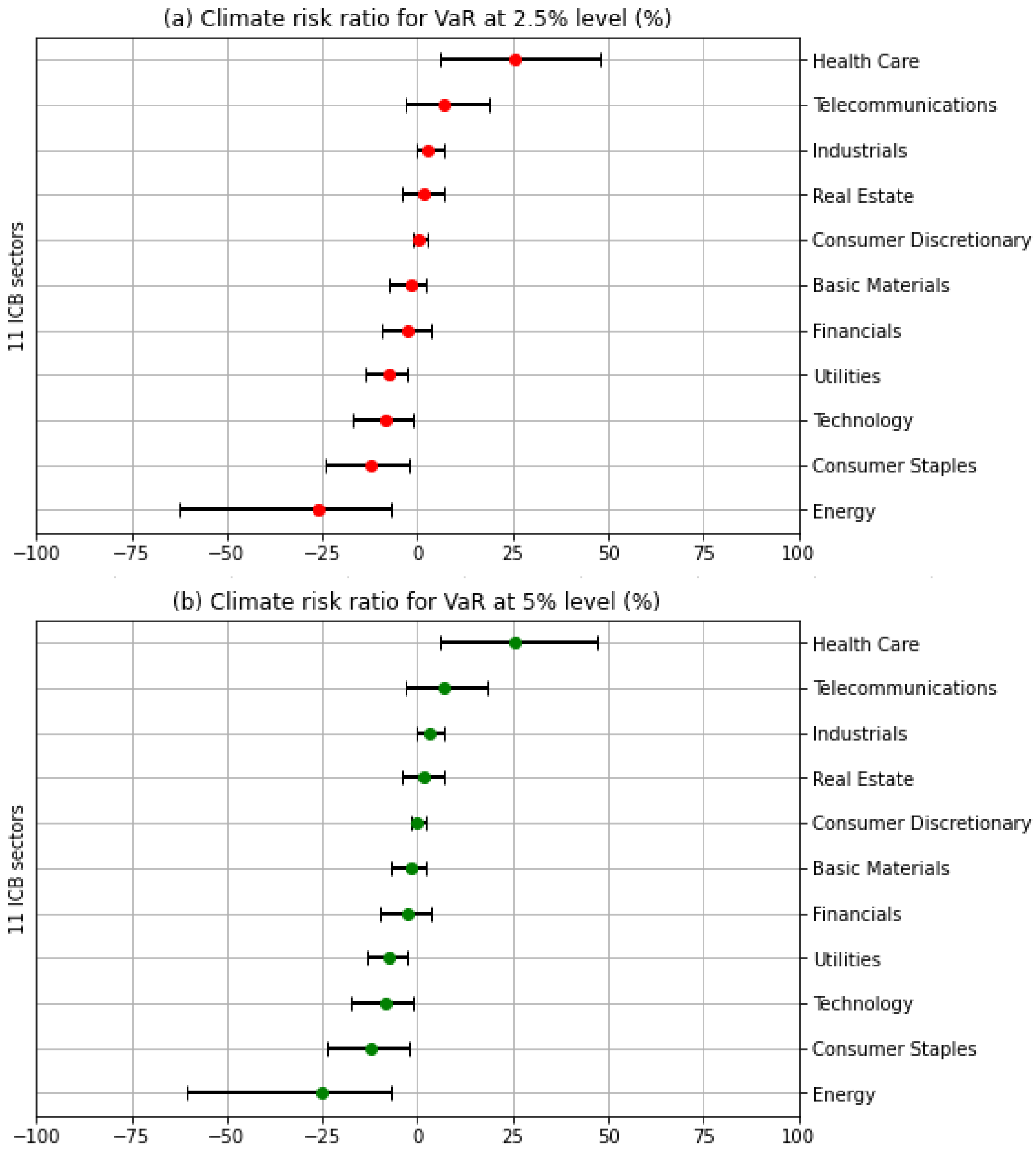
Table 3.5.7: Climate risk ratios and ratio rankings for VaR at 1% level

	Climate risk ratio			Rank		
Sector	GARCH-SKT	GJR-GARCH-SKT	GARCH-FZ	GARCH-SKT	GJR-GARCH-SKT	GARCH-FZ
Basic Materials	-1.798	-0.502	1.266	6	6	7
Consumer Discretionary	1.026	1.951	1.265	7	7	6
Consumer Staples	-12.382	-16.257	-11.202	2	2	2
Energy	-26.996	-28.556	-27.457	1	1	1
Financials	-1.896	-1.355	-3.521	5	5	5
Health Care	26.499	27.649	20.652	11	11	11
Industrials	2.585	4.128	2.018	9	9	8
Real Estate	1.879	1.964	2.149	8	8	9
Technology	-8.086	-8.250	-7.669	3	3	4
Telecommunications	7.141	8.699	12.404	10	10	10
Utilities	-7.776	-6.516	-8.355	4	4	3

Note: This table presents the average climate risk ratios (in percentage) and the rankings for 11 sectors (the model with the lowest ratio is ranked 1 and the model with the highest ratio is ranked 11) based on the climate risk ratio for VaR estimates at 1% level from January 2003 to December 2019 for 3 risk model specifications. The negative (positive) ratio refers to a reduction (increase) in the total risk due to environmental scores. The GARCH-SKT, GJR-GARCH-SKT, and GARCH-FZ correspond to the GARCH model with skewed t distribution, the GJR-GARCH model with skewed t distribution, and the GARCH model estimated using the FZ0 loss function of Fissler and Ziegel (2016), respectively.

skewed t distribution, as dependent variables in Eq. (3.2.5).[8] Figure 3.5.4 presents the summary of the climate risk ratio for VaR at 2.5% and 5% levels for the 11 sectors previously considered. Figure 3.4.3 and Figure 3.5.4 are similar, in that the ranking position of all sectors corresponds between the two figures. The 5% (95%) quantile of the climate risk ratio for companies in the Energy (Health Care) sector at 1% risk level is on average -65.293 (49.956), and at the 5% risk level, it is -60.268 (46.893). By shifting 1% risk levels to less extreme risk levels, the influence of environmental scores on downside risk is reduced, with the exception of companies in the Financials, Industrials, and Technology sectors, which have 5% risk levels on average more impacted by the companies' environmental scores.

Figure 3.5.4: Summary statistics of the climate risk ratio for VaR



Note: This figure presents the summary statistics of the climate risk ratio (in percentages) for VaR at 2.5% (a) and 5% (b) levels for 11 sectors. The left and right boundaries of the error bars are the 5 percent and 95 percent quantiles of the ratio, while the coloured marker represents the mean value. The sectors in both panels are ordered in descending order of the climate risk ratio

3.6 Conclusion

In this study, we propose new measures of climate downside risk that reveal to what extent the firm-level environmental scores influence the downside risk of the firms. We reveal the statistically significant negative relationship between stock returns and environmental scores at low quantiles. We employ the Emission score, Innovation score, and Resource Use score under the environmental pillar of the Thomson Reuters ASSET4 ESG scores as proxies for environmental information pertinent to the downside risk of the firms in various sectors. Our definitions of climate VaR and ES capture the market risk components associated to climate risk. We document that there is heterogeneity in the sensitivity of the firm-level risk to environmental scores. Our framework shows that firms in some sectors, notably Energy and Utilities, can reduce their downside risk by improving their firms' environmental scores, while for companies in sectors such as Health Care, improving the environmental scores is not cost-effective. These results are consistent with various risk assessments and levels of risk. These findings have important implications for investors and business managers to capture sensitivities to climate-related risk factors. Future research could consider a more nuanced decomposition of climate risk, in addition to the investigation of the relationship between downside risks and physical risk factors (e.g. rising sea levels or hurricane-prone regions).

Appendices

3.A Additional results

Table 3.A.1 provides the average correlations between environmental scores and control variables. Columns 1–3 show the correlation between the Emission, Innovation, and Resource Use scores. The correlation between the scores for Emission, Innovation, and Resource Use is shown in columns 1 through 3. The correlation between the Resource Use score and the Emission Score is 0.391, which is the greatest among the three types of scores. Also, we find that the environmental scores and control variables in our study are not highly correlated.

Table 3.A.1: Correlations of the environmental scores and control variables

Variable	*Emission*	*Innovation*	*Resource*	*Size*	*M/B*	*ROE*	*Leverage*	*Investment*
Emission	1.000							
Innovation	0.076	1.000						
Resource	0.391	0.100	1.000					
Size	0.116	0.044	0.141	1.000				
M/B	0.022	0.005	0.043	0.660	1.000			
ROE	0.008	0.003	0.012	0.135	0.150	1.000		
Leverage	0.098	0.038	0.102	-0.027	0.198	-0.043	1.000	
Investment	0.037	0.037	0.044	0.106	-0.004	0.021	-0.040	1.000

Note: This table shows the average firm pairwise correlations between the Emission score, Innovation score, Resource Use score, and the control variables during the sample period from January 2003 to December 2019. The correlation matrix is computed using data with monthly frequency.

Table 3.A.2 displays the sector average climate risk ratios (in percentages) and

rankings for ES estimated from three different models, specifically the GARCH model with skewed t distribution, the GJR-GARCH model with skewed t distribution, and the GARCH model estimated using the FZ0 loss function. We anticipate ES to be more negative than the corresponding VaR since, according to Eq.(3.2.3), ES is the expected value of losses below the VaR. Similar to Table 3.5.7, Energy and Health Care are the sectors that are most influenced by environmental scores, with climate risk ratios ranging from -27.940 to -31.096 for the former and from 22.451 to 28.81 for the latter. According to the ranking of climate risk ratios, the results of both alternative models are similar. Compared with the baseline model, the ranking of 7 sectors is consistent.

Table 3.A.2: Climate risk ratios and the ratio rankings for ES at 1% level

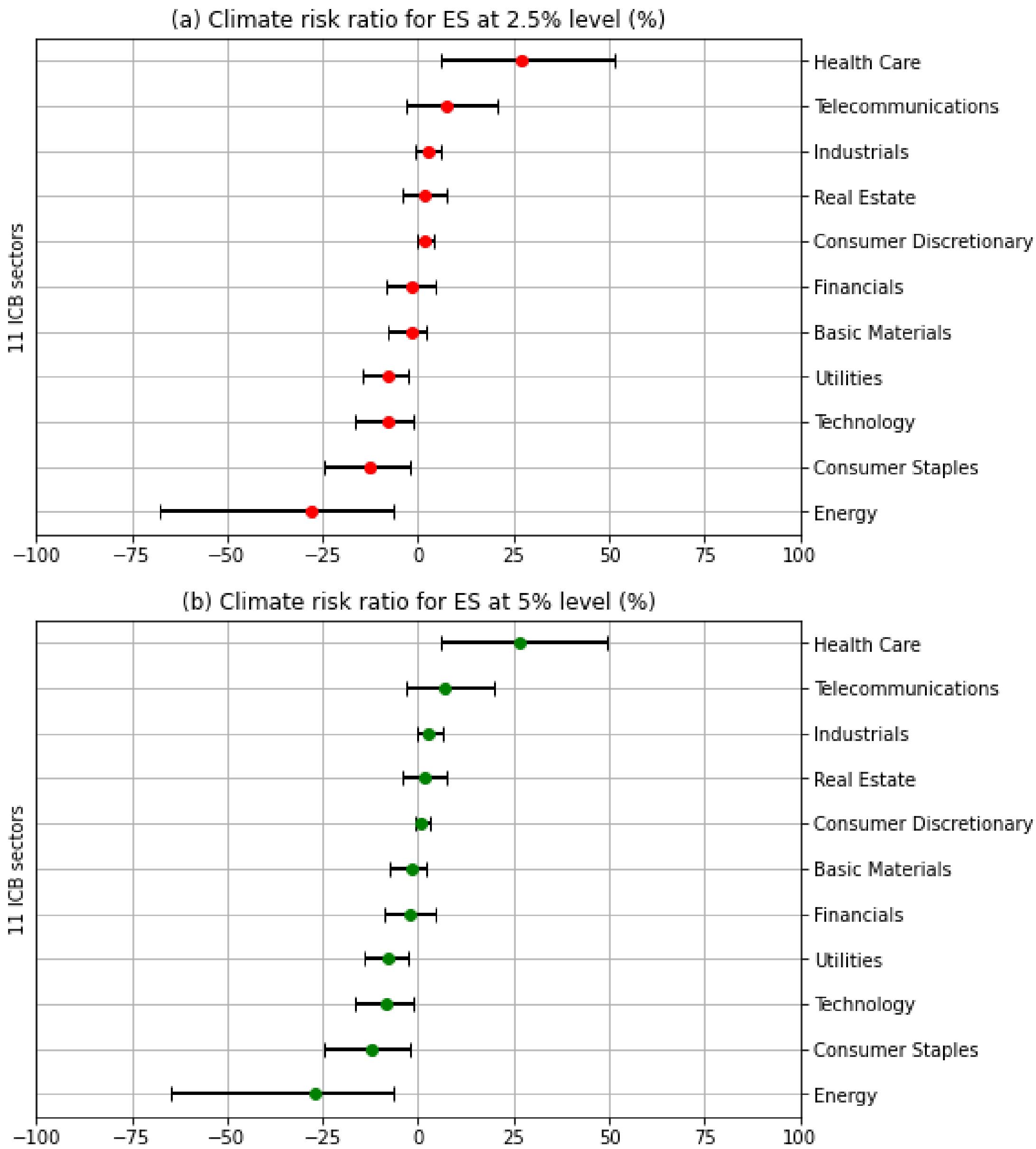
	Climate risk ratio			Ranking		
	GARCH-SKT	GJR-GARCH-SKT	GARCH-FZ	GARCH-SKT	GJR-GARCH-SKT	GARCH-FZ
Basic Materials	-1.947	-0.689	1.011	5	5	6
Consumer Discretionary	2.249	3.206	3.749	9	8	9
Consumer Staples	-12.771	-16.825	-11.800	2	2	2
Energy	-29.397	-31.096	-27.940	1	1	1
Financials	-0.805	-0.553	-3.300	6	6	5
Health Care	27.671	28.923	22.451	11	11	11
Industrials	2.223	3.751	1.396	8	9	7
Real Estate	1.900	2.002	2.473	7	7	8
Technology	-7.954	-8.177	-6.898	4	3	4
Telecommunications	7.382	8.756	14.981	10	10	10
Utilities	-8.314	-6.892	-9.181	3	4	3

Note: This table presents the average climate risk ratios and the rankings for 11 sectors (the model with the lowest ratio is ranked 1 and the model with the highest ratio is ranked 11) based on the climate risk ratio for ES estimates at 1% level from January 2003 to December 2019 across three risk model specifications. The negative (positive) ratio refers to a reduction (increase) in the total risk due to environmental scores. GARCH-SKT, GJR-GARCH-SKT, and GARCH-FZ correspond to the GARCH model with skewed t distribution, the GJR-GARCH model with skewed t distribution, and the GARCH model estimated with the FZ0 loss function from Fissler and Ziegel (2016), respectively.

Figure 3.A.1 presents the summary of the climate risk ratio for ES (in percentages) at 2.5% and 5% levels for 11 sectors. In terms of the mean value, we observe a similar pattern at 1% (in Figure 3.2.7), 2.5%, and 5% levels, where the

climate risk ratio ranges from around -29 to 27.5 across the three levels. When the risk level changes from 1% to 5%, the 95% quantile of the ratio for the Health Care sector progressively declines, but the 5% quantile of the ratio for the Energy sector gradually rises. Health Care, Telecommunications, Consumer Staples, and Energy hold the same ranking position when compared with the climate risk ratio for VaR.

Figure 3.A.1: Summary statistics of the climate risk ratio for ES



Note: This figure presents the summary statistics of the climate risk ratio (in percentages) for ES at 2.5% (a) and 5% (b) levels for 11 sectors. For each sector, the left and right limits of the error bar are the 5 percent and 95 percent quantiles of the ratio, while the coloured marker represents the mean value. The sectors in both panels are ordered in descending order of the climate risk ratio

Notes

[1]The three scores are not correlated with market returns so our measures of climate risk are not capturing market risks, according to our analysis.

[2]Following the approach in Bolton and Kacperczyk (2021), we run these regressions for firm-months observations. The firm-level control variables are updated quarterly, so in our regressions, we use the most recent observation for these variables. The emission score variables are updated annually, and for these as well we use the most recent observations in our regressions.

[3]The environmental scores are between 0 and 100, and the risk is typically expressed as a negative number.

[4]The data frequency is monthly, and control variables are lagged one-month to address the endogeneity issue.

[5]We follow Koenker (2004) in the specification of Eq.(3.2.11), because a large number of individual fixed effects can significantly inflate the variability of estimates of other covariate effects. Shrinkage of these individual effects toward a common value can help to modify this inflation effect.

[6]The correlations of the environmental scores and control variables are reported in the Supplemental Appendix.

[7]We have not investigated reverse causality and the affect of possibly missing variables; this might affect our results and we leave this for future research.

[8]Analogous results for ES are available in the Supplemental Appendix.

Chapter 4

Environmental performance and Credit Ratings: A Transatlantic study

4.1 Introduction

In this study, we investigate whether high environmental performances contribute to improvements in the US and European firms' credit ratings and how the influence of corporate environmental factors differs between firms in the two regions. This question is of particular significance given the increasing attention to companies' environmental performance over time (Klassen and McLaughlin, 1996; Bauer and Hann, 2010; Dyck et al., 2019; Christensen et al., 2021; Trinh et al., 2023), and uncovers the environmental determinants of credit ratings in the world's two largest economies. The implications are substantial: even minor improvements in credit ratings can result in reduced debt costs, fewer debt issues, and increased capital investment (Tang, 2009; Baghai et al., 2014).

This paper provides an important update on this research question, as Environmental, Social and Governance (ESG) and Corporate Social Responsibility (CSR) was shown to be an irrelevant factor for corporate bond pricing in the past (Menz, 2010).[1] However, this relation might have changed given that credit rating agencies recently incorporate environmental information into their assessments of debt issuers' creditworthiness. This is due to the growing importance of the firms' environmental and social activities, which impact both their financial and non-financial attributes, such as management strength and long-term sustainability (Attig et al., 2013). Given the differing perceptions and regulatory requirements of ESG/CSR between the US and EU, we further posit that the influence of environmental performance on credit ratings differs across these two regions.

To empirically examine the effects of firms' environmental factors on their credit ratings, we use the environmental pillar of the ESG ratings provided by Thomson Reuters ASSET4 ESG database as a measure of the company's environmental performance. In this chapter, we are looking closely at the whole environmental pillar of the ESG score because we want to investigate the association between overall environmental performance and credit ratings. Our rating sample includes long-term foreign currency issuer ratings issued by S&P, Moody's, and Fitch. We employ two methodologies to transform credit rating into scores: (1) we combine credit ratings, watches and outlooks together into numerical values ranging from 0 to 58 for the OLS model, and (2) we only consider credit rating signals and transform rating notches into ordinal numbers from 1 to 20 for the ordinal logit model.

Our findings suggest that rating agencies tend to grant firms with higher environmental scores better credit ratings. Moreover, we find that the impact of environmental performances on firms' ratings differs between the US and EU. This can be partially explained by the differences in the level of environmental performance in the two regions, in line with Cai et al. (2016) and Liang and Renneboog (2017). The EU's more strict ESG/CSR regulations result in better environmental performance of their firms (Christensen et al., 2021), whereas environmental or social performance disclosure is optional in the US. Thus, credit rating agencies are likely to evaluate the implications of an increase in environmental performance differently across these two regions. For instance, US firms that improve their environmental scores can be perceived as more proactive due to their country's less stringent environmental policy (Chava, 2014), and such voluntary improvements may be rewarded. In contrast, the EU has the norm of a high-level environmental consciousness, and thus an additional improvement in EU firms' environmental performance may have smaller benefits on their credit ratings, especially considering strict penalties for non-compliance (Paris Agreement, 2015).

Our first contribution is to investigate how improvements in firms' environmental performance affect their credit ratings. Previous studies have examined the factors that influence credit rating in several areas. A few of these focus on CSR and corporate social performance (CSP) in the US (Attig et al., 2013; Oikonomou et al., 2014; Ge and Liu, 2015) and in the EU (Menz, 2010; Stellner et al., 2015). Some studies document the correlation between firms' credit ratings and environmental performance in the US (Bauer and Hann, 2010; Seltzer et al., 2022; Safiullah et al., 2021). We extend this line of research by investigating the relation between firm-level environmental scores and credit ratings with a more comprehensive credit rating measure and a more recent dataset in both US and EU.

Our second contribution is to provide insights on the regional differences between the US and the EU regarding the impact of firms' environmental performance on credit ratings. Cai et al. (2016), Liang and Renneboog (2017), and Christensen et al. (2021) find that the ESG/CSR level is generally higher in the EU than in the US. Inspired by this line of research, our findings enrich the existing literature by suggesting that regional environmental norms also affect the influence of firms' environmental performance on their credit ratings.

The remainder of this paper is organised as follows. Section 2 presents the hypothesis development. Section 3 elaborates on the construction of numerical credit ratings and the baseline model. Section 4 outlines the data and provides summary statistics. Sections 5 and 6 present the main results and endogeneity tests, respectively. Section 7 offers additional robustness tests, and section 8 concludes.

4.2 Hypothesis development

In view of stakeholder theory (Freeman, 1984), firms that demonstrate high social responsibility are more likely to establish positive relationships with various stakeholders, including employees, consumers, suppliers, investors, and regulators

(Waddock and Graves, 1997; Fombrun and Shanley, 1990). Good relationship management helps to create valuable intangible assets, such as higher customer loyalty, and the ability to attract and retain high-quality employees (Turban and Greening, 1997; Greening and Turban, 2000). From a resource-based perspective, firms with superior environmental performance can effectively utilize their internal resources for sustainability initiatives and reduce the risk of financial distress (Barney, 1991; Attig et al., 2013). Prior studies find a positive correlation between individual ESG dimensions and firms' financial performance (Clarkson et al., 2011; Gompers et al., 2003). In particular, Gompers et al. (2003) demonstrate that firms with robust corporate governance experience higher firm value, profits, sales growth, lower capital expenditures, and fewer corporate acquisitions.

Firms' environmental performance is a crucial criterion for their interactions with various stakeholders. In the US, firms' environmental implications (e.g., the emission of toxic chemicals and hazardous waste) are already regulated by the government (Chava, 2014). By contrast, the disclosure of ESG/CSR performance is not mandatory (Christensen et al., 2021), which puts emphasis on the environmental pillar of the ESG rating. Additionally, firms with a strong environmental profile are more like to have superior fundamentals (Klassen and McLaughlin, 1996; Russo and Fouts, 1997; Konar and Cohen, 2001; Clarkson et al., 2011), more external resources from institutional investors (Chava, 2014; Fernando et al., 2017; Dyck et al., 2019; Tang and Zhang, 2020), lower cost of capital (Ng and Rezaee, 2015; Sharfman and Fernando, 2008; Chava, 2014; Bauer and Hann, 2010), and low risks (Seltzer et al., 2022; Hoepner et al., 2018; Jagannathan et al., 2018; Ilhan et al., 2021; Feldman et al., 1997).

Noting the increased attention on ESG/CSR performance in developed countries (Cai et al., 2016), we posit that in both the US and the EU, the credit rating agencies, recognizing the long-term sustainability and reduced risk associated with strong environmental practices, are more inclined to assign higher credit ratings to firms with enhanced environmental performance. This trend reflects a growing awareness that effective environmental management is not only a marker of corporate responsibility but also a key contributor to financial stability and resilience, making such firms more creditworthy. For this reason, we propose the following hypothesis:

H1. In both the US and the EU, an enhancement in a firm's environmental performance contributes to an improvement in its credit rating.

Although we expect that the positive impact of a firm's environmental improvement on its creditworthiness exists in both economies, the magnitude of this effect may differ between the US and European firms due to differentiated regulatory requirements. According to the empirical results of Cai et al. (2016), developed countries show higher Intangible Value Assessment (IVA) ratings, and most of the EU countries in their sample present greater ESG/CSR ratings than the US. Liang and Renneboog (2017) also employ the IVA ratings to show that most of the EU countries have higher ESG/CSR ratings and are more willing to follow environmental regulation or policy on CO2 emission than the US. The EU's Non-Financial Reporting Directive requires large companies with over 500 employees to include non-financial and diversity information in their management reports (European Union, 2014). Additionally, The EU is currently reviewing this directive and considering ways to strengthen the disclosure requirement, for example, by imposing additional audit standards (European Commission, 2020). In contrast, US firms publish CSR-related information either on a voluntary basis or when disclosure is material to investors under existing securities law (Christensen et al., 2021).

Additionally, EU countries are more advanced with environment protection policies. Before the Paris Agreement, the approach of the EU to deal with environmental issues has been based on binding targets among member states, implementing these through a common legislative framework including the EU Emissions Trading System (ETS).[2] On the contrary, the US has not passed any major climate change legislation in the last ten years. Even if both US and EU are parties to the United Nations Framework Convention on Climate Change (UNFCCC), EU countries have committed to emission reductions under the Kyoto Protocol, while the US did not ratify the protocol (Paris Agreement, 2015). Based on evidence from these two aspects, it can be concluded that the EU has a more develop environmental legislative framework than the US, which has lead to a higher environmental performance of firms in the EU compared to the US.

Considering the above-mentioned aspects, firms demonstrating superior environmental performance are likely to be rewarded by rating agencies as recognition of their proactive environmental efforts in the US. By contrast, since EU has a more stringent climate-related and environmental regulation framework, rating agencies may already have priced high environmental performance in their assessment of firms' creditworthiness. Given the distinct regulatory environments and differing degrees of environmental performance emphasis in the US and EU, we posit further improvements in European companies' environmental factors may not lead to substantial increases in credit ratings. This expectation leads to our second hypothesis:

H2. Firms' credit rating improvements associated with enhancements in their environmental performance are more pronounced in the US than in the EU.

4.3 Data

In this section, we illustrate the data sample and summary statistics. Our sample consists of firm-level environmental performance (measured by environmental scores from the Thomson Reuters ASSET4 ESG database) and long-term foreign-currency credit ratings issued by the three leading credit rating agencies (CRAs), namely Standard & Poor's, Moody's, and Fitch.

4.3.1 Sample construction

The credit rating sample is extracted from Bloomberg and contains three types of rating signals for all non-financial firms in the US and the EU: long-term foreign currency issuer ratings, credit watches and outlooks. The rating signals are issued by one of the three leading CRAs in the period from January 2003 to December 2022. According to Alsakka and Ap Gwilym (2013) and Alsakka et al. (2014), issuer ratings are transformed into numerical values according to a 20-point scale. Based on the numerical rating scale, upgrades (downgrades) are identified if the numerical current rating is higher (lower) than the previous one. Next, we consider credit watches and outlooks as additional rating signals. Positive (negative) watch signals, which by definition consist of placements on a rating agency's positive (negative) watch list, are either solo or combined signals. The former are identified as 'stand-alone' watch list placements, while the latter are watch signals accompanied by the same agency's rating changes. Positive (negative) outlook signals are additions to positive (negative) outlook lists for the countries with stable outlooks or no outlook announcement in advance. Similarly, outlook signals can also be solo or combined with rating changes.

In order to differentiate between solo and combined rating signals in a precise way, it is necessary to introduce a more powerful rating scale which fully takes the differences between solo and combined rating signals into consideration. For this purpose, the initial transformation based on a 20-point scale is extended to a 58-point system in line with Ferreira and Gama (2007) and Alsakka and Ap Gwilym (2013). The new rating scale is named as comprehensive credit rating (CCR) scale by prior literature. The CCR incorporates ratings, watch and outlook signals simultaneously in a new scale as follows: AAA/Aaa = 58, AA+/Aa1 = 55, AA/Aa2 = 52, ..., CCC-/Caa3 = 4, CC/Ca, SD-D/C = 1. In addition, "+2" ("-2") is adjusted for positive (negative) watch signal, while "+1" ("-1") is adjusted for positive (negative) outlook signal and "0" for stable outlook and no watch/outlook assignments.

We source our data of firms' environmental performance from the Thomson Reuters ASSET4 ESG database. This database gathers information from various sources such as annual reports, corporate sustainability reports, nongovernmental organizations, and news media, focusing on large, publicly traded companies across more than 45 countries on an annual basis. According to Thomson Reuters, the selection of data items aims at optimizing factors like company coverage, timeliness, data availability, quality, and perceived relevance for investors. To assess firms' environmental commitment, ASSET4 issues scores to three key areas: Emission Reduction, Product Innovation, and Resource Reduction. These environmental scores range from 0 to 100, where higher scores indicate better environmental performance.

The original frequency of both firm-level fundamentals and environmental scores is yearly. As company credit ratings can be updated multiple times per year, in order to align the frequency of firm-level variables with that of rating signals, we set up a monthly panel to include the two data sources. As a result, our initial sample contains 523,522 firm-month level observations of 1734 firms. We eliminate 138,044 firm-month observations that are missing the environmental scores and 37,548 firm-month observations that are missing financial statement data from Compustat. Our final sample consists of 347,930 observations of 1486 firms.[3]

Table 4.3.1 presents the sample distribution by credit rating agency, industry, and year. S&P is the most widely used credit rating agency in both subsamples. From the point of view of industry representation, Consumer Discretionary and Industrials are most present in both US and EU samples. Overall, the number of observations has risen gradually over the sample period, with a slight decrease in the final year, likely due to incomplete data availability for that particular year.

4.3.2 Summary statistics

Table 4.3.2 (Panel A) presents the descriptive statistics for all variables employed in our empirical analyses. The mean $RATING\_20$ score sits just under 11 (equiv-

Table 4.3.1: Sample description by Agency, Industry, and Year.

Panel A: Composition by Agency			Panel B: Composition by Industry			Panel C: Composition by Year		
	Observations			Observations			Observations	
Agency	EU	US	Industry	EU	US	Year	EU	US
Fitch	25,638	54,109	Real Estate	102	6140	2003	1954	2947
Moody's	30,973	77,273	Telecommunications	10,775	7459	2004	2327	3418
S&P	48,060	111,877	Technology	2259	21,352	2005	3060	4779
			Energy	4895	19,717	2006	3987	5573
			Health Care	6327	19,817	2007	4226	6036
			Basic Materials	9672	16,712	2008	4429	6763
			Consumer Staples	9804	19,935	2009	4609	8516
			Utilities	14,288	21,612	2010	4758	9288
			Industrials	24,897	54,504	2011	5010	9830
			Consumer Discretionary	22,483	57,946	2012	5163	10,396
						2013	5204	10,649
						2014	5516	11,350
						2015	5860	12,190
						2016	6034	15,947
						2017	6204	18,373
						2018	6670	20,033
						2019	7125	20,834
						2020	7425	22,023
						2021	7946	23,943
						2022	7164	20,371
Total	104,671	243,259		104,671	243,259		104,671	243,259
Firms	472	1014		472	1014		472	1014

Notes: This table presents the number of observations by agency, industry, and year in Panels A, B, and C, respectively. This sample covers the long-term issuer credit ratings from S&P, Moody's, and Fitch, 10 ICB industries, and the period ranging from January 2003 to December 2022.

alent to a BBB– rating), with a standard deviation of around 3 and an interquartile range of 4.[4] Notably, the statistics for $RATING\_58$ are roughly triple that of $RATING\_20$. The average environmental score is around 45, with a standard deviation of about 29, which suggests a wide range of environmental performance across firms. On average, debt leverage is around 34% of the total assets of the firms, while the mean ROA is 4.23%. The mean $SIZE$ is around 9, indicating that our sample firms are generally large. The mean of $BIG4$ is 0.9489, which demonstrates that the majority of firms in our sample are audited by one of the Big 4 audit firms.

Panel B of Table 4.3.2 provides a comparative summary of the statistics between firms in the US and the EU. Credit ratings appear to be slightly higher for EU firms than for their US counterparts. EU firms also show higher environmental scores, capital intensity, profit margin, and larger firm size. By contrast, they present lower losses, lower leverage, and a smaller standard deviation of operational cash flow and ROA compared to US firms. These findings suggest that, on average, EU firms demonstrate stronger financial and environmental performance compared to US firms.

In Figure 4.3.1, we illustrate the average environmental scores by firms in different industries and years for the US and EU samples. Considerable variation can be observed in the environmental scores across different industries and years. Figure 4.3.1a displays the average environmental scores for various industries. It is to be noted that environmental scores are consistently higher for firms in the EU sample, and this trend persists even in industries known to have high emissions, such as Energy and Utilities. In the EU sample, these industries demonstrate

Table 4.3.2: Summary Statistics for the US and EU samples

Panel A: Full Sample statistics (N = 347,930)

Variables	Mean	S.D.	Min	Q1	Median	Q3	Max
RATING_58	30.1242	9.7803	0.0000	23.0000	31.0000	37.0000	58.0000
RATING_20	10.9990	3.1713	1.0000	9.0000	11.0000	13.0000	20.0000
ENV	45.8616	29.3326	0.0000	20.0000	48.8653	71.6250	99.1667
SIZE	9.2748	1.3743	4.3633	8.3168	9.1587	10.2401	14.1525
ROA	0.0423	0.0670	-0.2479	0.0173	0.0412	0.0740	0.2346
LOSS	0.0738	0.2615	0.0000	0.0000	0.0000	0.0000	1.0000
LEV	0.3404	0.1772	0.0021	0.2178	0.3215	0.4383	1.0013
INT_COV	13.8769	22.2931	-3.7178	4.7070	7.9797	13.9413	168.0000
CAP_INTEN	0.6003	0.4140	0.0045	0.2490	0.5383	0.8975	1.9075
BIG4	0.9489	0.2201	0.0000	1.0000	1.0000	1.0000	1.0000
CFO_STD	0.0296	0.0237	0.0033	0.0135	0.0225	0.0377	0.1357
ROA_STD	0.0373	0.0473	0.0019	0.0117	0.0223	0.0427	0.3302
MARGIN	0.2055	0.1455	-0.1643	0.1104	0.1762	0.2777	0.7180

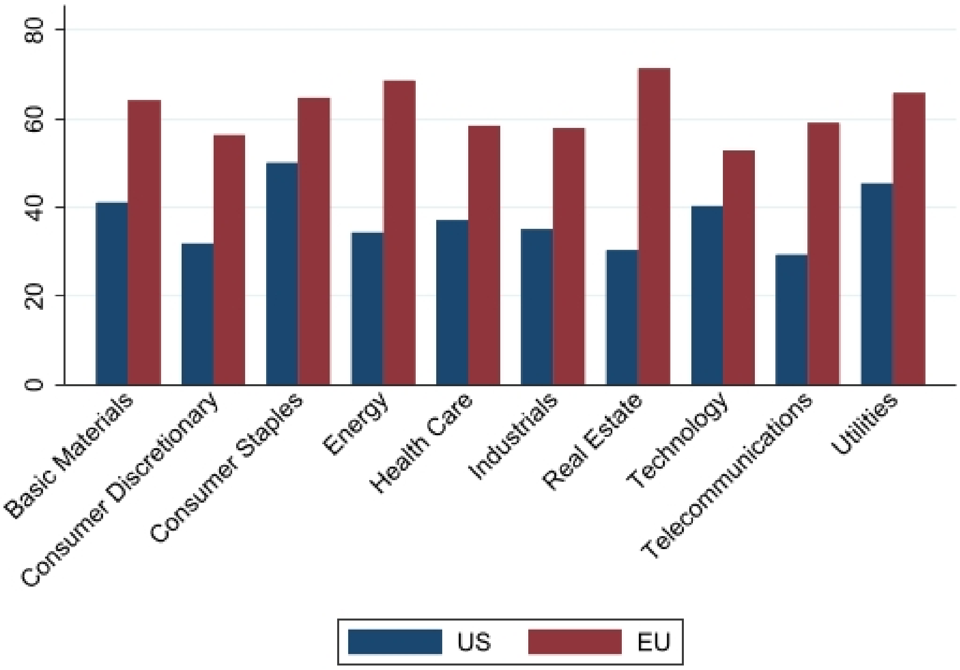
Panel B: Descriptive statistics of firm variables in the two regions

Variables	US (N = 243,259)			EU (N = 104,671)		
	Mean	Median	S.D.	Mean	Median	S.D.
RATING_58	29.3796	31.0000	10.0124	31.8546	34.0000	8.9831
RATING_20	10.7536	11.0000	3.2426	11.5694	12.0000	2.9203
ENV	39.1938	38.8154	28.6842	61.3580	66.9823	24.5840
SIZE	9.1420	9.0339	1.3380	9.5836	9.5435	1.4076
ROA	0.0449	0.0446	0.0704	0.0364	0.0358	0.0579
LOSS	0.0771	0.0000	0.2667	0.0663	0.0000	0.2487
LEV	0.3516	0.3309	0.1835	0.3144	0.3009	0.1586
INT_COV	14.1841	7.8934	23.2267	13.1628	8.1413	19.9372
CAP_INTEN	0.5877	0.5090	0.4154	0.6294	0.5989	0.4093
BIG4	0.9688	1.0000	0.1739	0.9028	1.0000	0.2962
CFO_STD	0.0313	0.0241	0.0245	0.0258	0.0193	0.0213
ROA_STD	0.0399	0.0236	0.0504	0.0313	0.0197	0.0383
MARGIN	0.2030	0.1790	0.1402	0.2112	0.1710	0.1568

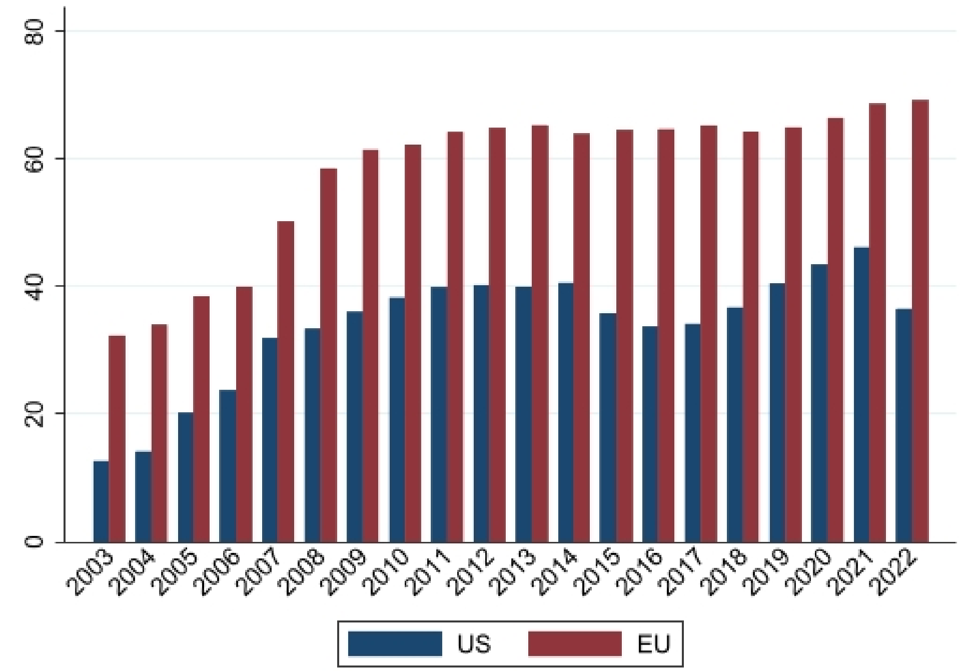
Notes: Panel A presents full sample descriptive statistics. Panel B presents the sample descriptive statistics for the two regions, the US and the EU.

Figure 4.3.1: Average environmental scores for the US and the EU firms

(a) Average environmental scores by industry



(b) Average environmental scores by year



This figure shows equal-weighted average environmental scores for US and EU firms. Figure (a) demonstrate the average environmental score of firms in each of the ICB industries, while figure (b) shows the average environmental scores of firms ranging from 2003 to 2022.

relatively high environmental scores over 60. Figure 4.3.1b shows the fluctuations in the environmental score throughout the sample period, spanning from 2003 to 2022. The disparities in environmental scores between US and EU firms persist on a year-to-year basis, as evidenced by the industry-level differences. EU firms consistently outperform their US counterparts in terms of environmental scores over the entire period. In summary, the differences in environmental scores displayed in Figure 4.3.1 suggest that EU firms tend to be more environmentally conscious compared to US firms, which is observable across industries and over time. These findings emphasize the substantial role that geographical location and industry characteristics may play in the environmental performance of firms.

The Pearson correlation matrix of the firm-level variables in the US and EU samples are reported in Table 4.3.3. The correlation coefficients between credit ratings and environmental scores are positive. The US sample correlation is around 0.37 while the EU sample presents a correlation of about 0.25. The results suggest that firms with higher environmental scores are likely to receive higher credit ratings and this positive relation might differ across the US and EU.

The Pearson correlation coefficients demonstrate that there are no extreme correlations between our control variables. To further test for multicollinearity issues, we investigate the variance inflation factors (VIFs). The average of the VIFs in our model is 1.45 (1.46) for the US (EU) sample, and none of the variables have VIFs greater than the critical value of 2.5 (Johnston et al., 2018).[5]

Table 4.3.3: Correlation matrix for the US and EU samples

Variables		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
RATING_58	(1)	1	**0.9628**	**0.2486**	**0.4691**	**0.2983**	**-0.2960**	**-0.1925**	**0.1311**	**0.0157**	**-0.0453**	**-0.2862**	**-0.3277**	**0.1753**
RATING_20	(2)	**0.9514**	1	**0.2702**	**0.5151**	**0.3016**	**-0.3058**	**-0.1929**	**0.1325**	**0.0326**	**-0.0331**	**-0.2972**	**-0.3354**	**0.1781**
ENV	(3)	**0.3722**	**0.4045**	1	**0.4728**	**0.0128**	**-0.0519**	**-0.0610**	**0.0261**	**0.0486**	**-0.0128**	**-0.2118**	**-0.2074**	**-0.0250**
SIZE	(4)	**0.5013**	**0.5434**	**0.5291**	1	**-0.0207**	**-0.0920**	**-0.1488**	**-0.0068**	**0.0350**	**-0.0196**	**-0.3041**	**-0.2546**	**0.0206**
ROA	(5)	**0.3771**	**0.3808**	**0.1354**	**0.0817**	1	**-0.4433**	**-0.1525**	**0.3703**	**-0.0647**	**0.0401**	**0.0154**	**-0.1554**	**0.2629**
LOSS	(6)	**-0.2957**	**-0.3141**	**-0.1003**	**-0.1346**	**-0.4875**	1	**0.1201**	**-0.1247**	**0.0282**	0.0044	**0.1353**	**0.2354**	**-0.1581**
LEV	(7)	**-0.3205**	**-0.3330**	**-0.0856**	**-0.1250**	**-0.1920**	**0.1651**	1	**-0.3783**	**0.1713**	**0.0304**	**0.0170**	**0.0761**	**0.4529**
INT_COV	(8)	**0.2117**	**0.2015**	**0.0600**	**0.0276**	**0.3607**	**-0.1373**	**-0.4379**	1	**-0.0500**	**-0.0309**	**0.0896**	**-0.0158**	**0.0433**
CAP_INTEN	(9)	**-0.0511**	**-0.0426**	**0.0690**	**0.0237**	**-0.1630**	**0.0718**	**0.1086**	**-0.1027**	1	**0.0716**	**-0.0618**	**-0.0572**	**0.2609**
BIG4	(10)	**0.2074**	**0.2056**	**0.1525**	**0.2175**	**0.1079**	**-0.0978**	**-0.0962**	**0.0437**	**-0.0569**	1	**0.0128**	**0.0304**	-0.0018
CFO_STD	(11)	**-0.2232**	**-0.2475**	**-0.1538**	**-0.2609**	**-0.0413**	**0.1590**	**0.0260**	**0.0568**	-0.0007	**-0.0545**	1	**0.4731**	**-0.0664**
ROA_STD	(12)	**-0.2999**	**-0.3290**	**-0.0963**	**-0.2034**	**-0.2619**	**0.2728**	**0.0959**	**-0.0413**	**0.1398**	**-0.1008**	**0.4676**	1	**-0.1096**
MARGIN	(13)	**0.1883**	**0.1944**	**0.0642**	**0.2286**	**0.3104**	**-0.1852**	**0.1205**	**0.0813**	**0.1918**	**-0.0281**	**-0.1476**	**-0.0864**	1

Notes: This table presents the Pearson correlation matrix of the firm-level variables. The numbers below (above) the diagonal are the Pearson correlation coefficients for US (EU) sample. Correlations significant at the 10% level are highlighted in bold.

4.4 Methodology

In our empirical tests, we employ OLS (ordinal logit) model for the numerical 58-point (ordinal 20-point) scale of credit ratings, controlling for several firm characteristics. The benefit of using the ordinal logit model is that it does not assume that each rating notch represents the same increase in a firm's rating; higher numbers are considered better ratings, but the exact magnitude of the rating is irrelevant. As our numerical rating scaled from 0 to 58 is linear as opposed to the regular numerical rating scaled from 1-20, which doesn't require such as assumption, there are benefits of employing the OLS estimation because it is more straightforward and it allows the analysis of economic significance based and it is consistent with the use of additional tests (Baghai et al., 2014). To account for possible correlations in the error terms, we adjust standard errors via firm-level clustering. The fundamental empirical specification in the baseline regression is given by the following equation:

$$RATING_{i,t} = \alpha + \beta ENV_{i,t-12} + \gamma X_{i,t-12} + \Lambda + \epsilon_{i,t}, \tag{4.4.1}$$

where $RATING_{i,t}$ constitutes the numerical conversion of the credit rating of firm i' at year-month t, with a higher value signifying superior creditworthiness, denoted as $RATING\_58$ or $RATING\_20$.[6] $ENV_{i,t-12}$, the key variable of interest, designates the environmental score from Thomson Reuters ASSET4 ESG database attributed to firm i at year-month $t-12$. If credit rating agencies consider a firm's environmental performance as one of the credit risk factors, we would expect β to be positive. The control variables in vector $X_{i,t-12}$ are also lagged by a year (twelve months) and are common throughout the different specifications. Λ are year-month, country, and industry fixed effects.

To isolate the effects of key variable of interest (environmental score), we control for a set of variables commonly used in literature of firm credit ratings (Cornaggia et al., 2017; Bhandari and Golden, 2021; Attig et al., 2013). These include: $SIZE$, the natural logarithm of total assets, expressed in millions of USD; ROA, the income before extraordinary items scaled by total assets; $LOSS$, an indicator variable set to 1 if income before extraordinary items is negative in the current and previous year, and 0 otherwise; LEV, total debt (long-term plus the portion of long-term debt in current liabilities) scaled by total assets; INT_COV, earnings before interest and taxes scaled by interest expense; CAP_INTEN, gross property, plant, and equipment scaled by total assets; $BIG4$, an indicator variable set to 1 if the auditor is a Big4 auditor, and 0 otherwise;[7] CFO_STD, the standard deviation of operating cash flows scaled by total assets for the previous 60 months; ROA_STD, the standard deviation of ROA for the previous 60 months; $MARGIN$, income before extraordinary items divided by sales. To mitigate the impact of outliers, we winsorize all continuous firm-level controls at the one and ninety-nine percentiles, except for $SIZE$, $LOSS$, and $BIG4$. Finally, we employ year-month, agency, industry, and country indicators to control for variations in ratings across different aspects.[8]

4.5 Empirical results

Table 4.5.4 reports the baseline regression results demonstrating the relation between environmental scores and credit ratings. In Column (1), the coefficient of ENV for the US firms is 0.0570, associated with a t-statistic of 6.50, signifying that the variable ENV is statistically significant at a 1% level. As for the EU sample (Column 2), the coefficient of ENV maintains its significance at 1% level, with a coefficient value of 0.0498. Columns (3) and (4) present results of the ordinal logit model, showing that the coefficients of ENV for both US and EU samples are notably positive and significant at 1% level, with a value of 0.0158 and 0.0117, respectively. These results suggest that the environmental score is a crucial determinant of credit ratings, for both US and EU firms. The economic impact of our empirical results is also significant. Under the OLS regression specification, one standard deviation increase in ENV is associated with a 1.6349 (0.0570×28.6842) increase in the 58 scaled credit ratings in the US, and a 1.2224 (0.0498×24.5840) increase for EU firms.

Results of the baseline regression by OLS and ordinal logit model confirmed our Hypothesis 1. Moreover, it should be noted that the difference in coefficients between the US and the EU indicates that this effect is more prominent for US firms. This finding aligns with our second hypothesis, which suggests that the credit rating benefits associated with improved environmental performance are indeed more pronounced in the US than in the EU.

The results on the control variables in the model are generally consistent with prior research (Cornaggia et al., 2017; Ashbaugh-Skaife et al., 2006; Bhandari and

Table 4.5.4: Baseline results for the US and EU samples

Panel A: Main results

	(1)	(2)	(3)	(4)
	US	EU	US	EU
	OLS		Ordinal Logit	
Variables	Dependent Variable = RATING_58		Dependent Variable = RATING_20	
ENV	0.0570***	0.0498***	0.0158***	0.0117***
	(0.0088)	(0.0142)	(0.0021)	(0.0038)
SIZE	2.1373***	2.1609***	0.7089***	0.7722***
	(0.2409)	(0.2850)	(0.0662)	(0.0958)
ROA	31.0905***	24.2721***	8.6754***	8.5351***
	(2.7146)	(4.2376)	(0.7648)	(1.3279)
LOSS	-1.5701***	-2.8409***	-0.4406***	-0.7144***
	(0.4887)	(0.5210)	(0.1168)	(0.1668)
LEV	-8.2153***	-8.9565***	-2.3536***	-2.7639***
	(1.3907)	(2.0326)	(0.3434)	(0.6339)
INT_COV	0.0178	0.0190	0.0060*	0.0101**
	(0.0124)	(0.0117)	(0.0033)	(0.0039)
CAP_INTEN	-0.7527	0.0523	0.0240	0.0651
	(0.6258)	(0.8323)	(0.1690)	(0.2539)
BIG4	2.2613***	-0.6900	0.4230***	-0.1714
	(0.8419)	(0.6830)	(0.1627)	(0.2170)
CFO_STD	-11.7620	-40.4239***	-3.1004	-9.6492***
	(7.2890)	(10.7059)	(1.9111)	(3.3716)
ROA_STD	-16.3137***	-26.4226***	-5.5213***	-7.2303***
	(3.3958)	(6.7155)	(0.9794)	(2.2292)
MARGIN	2.8696*	11.6559***	0.7609*	3.2536***
	(1.7003)	(2.3208)	(0.4448)	(0.7268)
Time F.E.	YES	YES	YES	YES
Agency F.E.	YES	YES	YES	YES
Industry F.E.	YES	YES	YES	YES
Country F.E.	NO	YES	NO	YES
Firm clustered	YES	YES	YES	YES
Observations	243,259	104,671	243,259	104,671
Adj. R^2/Pseudo R^2	0.484	0.512	0.157	0.169

Panel B: Marginal effects of the ordinal logit model

Rating	US			EU		
	Probability at 75th pct.E-Score High E-Score [E-Score = 63.855]	Probability at 25th pct. E-Score Low E-Score [E-Score = 11.416]	High - Low	Probability at 75th pct.E-Score High E-Score [E-Score = 80.640]	Probability at 25th pct. E-Score Low E-Score [E-Score =45.834]	High - Low
AAA (=20)	0.3411%	0.1508%	0.1902%	No Obs.	No Obs.	
AA+ (=19)	0.2271%	0.1022%	0.1248%	0.0462%	0.0309%	0.0153%
AA (=18)	0.7559%	0.3492%	0.4066%	0.6588%	0.4523%	0.2065%
AA- (=17)	1.5697%	0.7607%	0.8090%	2.0761%	1.5004%	0.5758%
A+ (=16)	3.5453%	1.8537%	1.6915%	5.0371%	3.8238%	1.2133%
A (=15)	8.5612%	5.0919%	3.4693%	5.9753%	4.7572%	1.2181%
A- (=14)	7.4822%	5.0674%	2.4148%	12.1848%	10.3313%	1.8536%
BBB+ (=13)	13.0778%	10.0972%	2.9806%	17.9299%	16.5893%	1.3406%
BBB (=12)	16.7046%	15.1272%	1.5774%	19.2936%	19.5677%	-0.2741%
BBB- (=11)	11.9538%	12.4810%	-0.5273%	12.6802%	13.8488%	-1.1686%
BB+ (=10)	8.4827%	9.8087%	-1.3261%	6.5641%	7.5257%	-0.9616%
BB (=9)	8.3198%	10.4968%	-2.1769%	5.6672%	6.7200%	-1.0528%
BB- (=8)	7.2974%	10.0442%	-2.7468%	3.7675%	4.5869%	-0.8194%
B+ (=7)	4.7368%	7.0399%	-2.3031%	2.3647%	2.9279%	-0.5632%
B (=6)	3.6591%	5.8233%	-2.1642%	2.3069%	2.8915%	-0.5847%
B- (=5)	1.7291%	2.9220%	-1.1929%	1.7529%	2.2238%	-0.4710%
CCC+ (=4)	0.6181%	1.0806%	-0.4626%	0.8624%	1.1142%	-0.2518%
CCC (=3)	0.3418%	0.6069%	-0.2652%	0.2439%	0.3213%	-0.0774%
CCC- (=2)	0.0976%	0.1746%	-0.0770%	0.1154%	0.1535%	-0.0381%
C/CC/D (=1)	0.4991%	0.9215%	-0.4223%	0.4730%	0.6337%	-0.1608%

Notes: This sample contains firm-month observations from January 2003 to December 2022, using Eq.(4.4.1) regression models. Numerically transformed long-term issuer ratings by S&P, Moody's, and Fitch are used, with $RATING\_58$ scaled from 0 to 58 and $RATING\_20$ scaled from 1 to 20 (4.3.1). The environmental score (ENV) is provided by Thomson Reuters ASSET4. All regressions include year-month, agency, and industry fixed effects; country fixed effects apply only to the EU sample. Reported significance is based on robust standard errors clustered at firm level. Significance at 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Panel A outlines the baseline model coefficients with OLS results in columns (1) and (2), and ordinal logit in (3) and (4). Panel B details the marginal effects from the ordinal logit regression reported in Panel B, displaying probabilities for various ratings at low (25th percentile) and high (75th percentile) environmental scores for companies in the US and EU samples.

Golden, 2021; Attig et al., 2013; Bonsall IV et al., 2017; Hossain et al., 2023). Specifically, accounting variables that capture financial risk, such as $SIZE$, ROA, INT_COV, and $MARGIN$ (LEV, CFO_STD, $LOSS$ and ROA_STD), are significantly positively (negatively) associated with credit ratings, and their signs are consistent across all model specifications. CAP_INTEN is positively significant under the ordinal logit regressions which is in line with the literature, but for OLS regressions the coefficient is significantly negative for US companies and insignificant for EU firms. Finally, the corporate governance proxy, $BIG4$, reduces managerial opportunistic behavior, which increases credit ratings for the US sample but decreases it for the EU sample. Panel B reports the probability of different ratings when the environmental score is at the 25th and 75th percentiles. Consistent with expectations, in both samples, the probability of higher ratings is higher when environmental score is high. However, when we compare the marginal effects on ratings between high scores and low scores, their difference is greater for the US than the EU sample for all rating grades, except for BBB- ratings. The greatest difference in the US sample is the probability of being rated A, with a value of 3.4693%, whilst in the EU sample the greatest difference is for the A- rating, 1.8536%.These results again prove Hypothesis 1. Table 4.5.4 presents the baseline results of our study, in which the estimation is potentially affected by endogeneity. In section 4.6, we discuss the estimation using instrumental variables which addresses the potential problem of endogeneity.

Another way to investigate the difference between US and EU is by using a dummy variable ($HIGH\_ENV$) which is equal to one if the firm's environmental score is above the median of the environmental score, and zero otherwise. We conduct OLS and ordinal logit regressions using $HIGH\_ENV$ as an alternative measure for the environmental score to test whether the difference between the two markets is significant for firms with higher/lower environmental scores. The results are reported in Table 4.5.5. For the OLS regression specification, we find that the coefficient estimate on $HIGH\_ENV$ is positive and statistically significant at 1% level, with a value of around 2.15 for the US and 1.22 for the EU sample. This means that the relation between credit ratings and environmental scores is stronger for the US sample than the EU sample, which is consistent with expectations. Also the coefficient estimate on $HIGH\_ENV$ for the US sample is higher than the coefficient estimate from for EU sample. In terms of the ordinal logit model specification, the coefficient for the US sample (0.5647) is twice as large as the one of the EU sample (0.2897). This provides further evidence for our Hypothesis 2 that firms with high environmental scores are more likely to have a higher credit rating, and the effect is more pronounced for US firms as compared to firms in the EU.[9]

A question that arises naturally is why the relation between credit ratings and environmental scores is stronger in the US than in the EU. First, we visually examine the link between credit ratings and environmental scores. We sort the credit ratings into four groups and compare them across the environmental score bins. Figure 4.5.2a depicts the average credit ratings by environmental score bins, for both markets. The figure clearly demonstrates that firms with lower environmental scores tend to have lower average credit ratings in both samples. However, when the environmental scores are below 50, EU firms, on average, have higher credit ratings than their US counterparts. In contrast, in the 50-

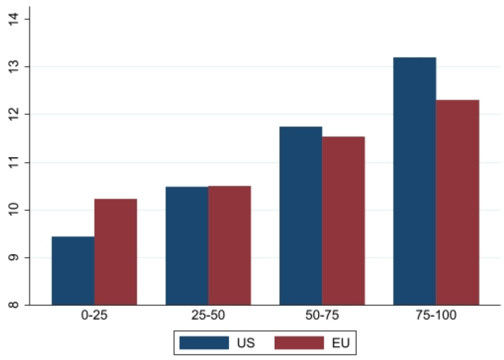
Table 4.5.5: Effect on high environmental score group and credit rating

	(1)	(2)	(3)	(4)
	US	EU	US	EU
	OLS		Ordinal Logit	
Variables	Dependent Variable = RATING_58		Dependent Variable = RATING_20	
HIGH_ENV	2.1494***	1.2218***	0.5647***	0.2897**
	(0.3558)	(0.4330)	(0.0881)	(0.1331)
SIZE	2.4114***	2.3964***	0.7833***	0.8256***
	(0.2360)	(0.2643)	(0.0645)	(0.0889)
ROA	32.5187***	25.4248***	9.0690***	8.8153***
	(2.7179)	(4.2758)	(0.7682)	(1.3366)
LOSS	-1.5251***	-2.7656***	-0.4243***	-0.6902***
	(0.4887)	(0.5329)	(0.1179)	(0.1680)
LEV	-8.4331***	-8.8714***	-2.3918***	-2.7321***
	(1.3916)	(2.0549)	(0.3409)	(0.6311)
INT_COV	0.0181	0.0196*	0.0061*	0.0102***
	(0.0126)	(0.0115)	(0.0033)	(0.0039)
CAP_INTEN	-0.4982	0.1927	0.0991	0.1004
	(0.6328)	(0.8245)	(0.1702)	(0.2511)
BIG4	2.5243***	-0.6608	0.5071***	-0.1508
	(0.8569)	(0.6743)	(0.1602)	(0.2139)
CFO_STD	-11.9825	-41.8765***	-3.1690*	-9.9325***
	(7.2976)	(10.8517)	(1.8918)	(3.3507)
ROA_STD	-15.6522***	-26.6548***	-5.2592***	-7.2299***
	(3.3785)	(6.8675)	(0.9611)	(2.2680)
MARGIN	2.3261	11.2268***	0.5716	3.1392***
	(1.7055)	(2.3403)	(0.4428)	(0.7366)
Time F.E.	YES	YES	YES	YES
Agency F.E.	YES	YES	YES	YES
Industry F.E.	YES	YES	YES	YES
Country F.E.	NO	YES	NO	YES
Firm Clustered	YES	YES	YES	YES
Observations	243,259	104,671	243,259	104,671
Adj. R^2/Pseudo R^2	0.477	0.506	0.153	0.168

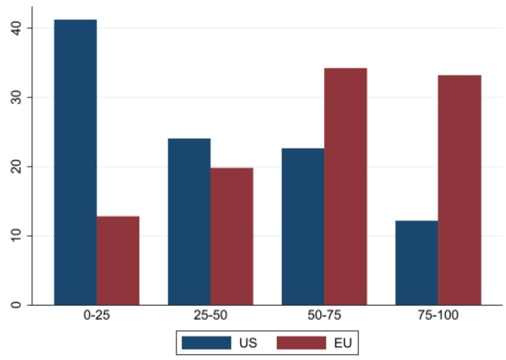
Notes: This table reports the results from regressions of credit ratings on high environmental score group. Columns (1) and (2) present results from the OLS specification, and columns (3) and (4) present those from the ordinal logit specification. We use the numerical transformation of domestic long-term issuer ratings by S&P, Moody's, and Fitch, increasing in credit quality. $RATING\_58$ is the credit rating scaled from 0 to 58, while $RATING\_20$ is scaled from 1 to 20, details in Section 4.3.1. $HIGH\_ENV$ equals one if the environmental score is above the median level of the environmental score, zero otherwise. All regressions include year-month, agency, and industry fixed effects, while country fixed effects are only employed in the EU sample. Reported significance is based on robust standard errors clustered at firm level. Significance at 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Figure 4.5.2: Average credit ratings and environmental scores in different levels

(a) Average ratings by levels of environmental score



(b) Histogram of environmental scores



This figure shows (a) equal-weighted average ratings for different categories of environmental scores and (b) the histogram of environmental scores, for US and EU samples.

100 range, US firms demonstrating superior environmental performance achieve better ratings than EU firms with similar environmental scores.

Figure 4.5.2b displays the histogram of environmental scores among firms in each group. In the US, more than 40% of observations are concentrated at levels with environmental scores below 25, and the number of observations decreases as the environmental score level increases. In contrast, the EU sample has only about 33% of observations with environmental scores below 50, while the remaining observations are evenly distributed across the two highest levels. This distribution in the EU sample is the other way around, with most of the observations having higher environmental scores. Interestingly, in the 0-25 environmental score bin, approximately 40% of observations have an environmental score of 0 in the US, compared to 22% in the EU.[10]

To further explore the observed patterns, we extend our baseline model to capture potential non-linear relation between environmental scores and credit ratings by adding a quadratic term for environmental scores, obtaining the following regression:

$$RATING_{i,t} = \alpha + \beta_1 ENV_{i,t-12} + \beta_2 ENV^2_{i,t-12} + \gamma X_{i,t-12} + \Lambda + \epsilon_{i,t}, \qquad (4.5.1)$$

where the dependent variables is the credit rating scaled from 0 to 58. ENV is the environmental score. We again include the same control variables and also control for fixed effects of agencies, industries, year-months, and countries (EU only). The results are reported in Table 4.5.6. In this setup we find that the relation between environmental performance and credit ratings is weakly significant in the US. However, for the EU sample, the coefficient of ENV is 0.1095 and statistically significant with a p-value below 0.01, which is twice as large as the coefficient of ENV from the baseline results (0.0498). The coefficient of ENV^2 is significantly negative at 10% level (-0.0006). This shows that there is a diminishing effect of the environmental score on credit ratings in the EU. In other words, the relation is strong and positive for low environmental scores, but it weakens for high values of the environmental score.

Fig. 4.5.3 presents the relation between the environmental score and the numerical transformation of credit ratings, ranging from 0 to 58, for both US and EU samples. In the US, the relation appears almost linear. In contrast, the EU depicts a decrease in marginal effects as ENV increases. Also, the relation between the two variables disappears for firms with an environmental score larger than about 80. The marginal impact on ratings spans from 0.1095 (evaluated at

Table 4.5.6: Results for non-linear relationship

	(1)	(2)
	US	EU
Variables	Dependent Variable = RATING_58	
ENV	0.0336†	0.1095***
	(0.0213)	(0.0388)
ENV^2	0.0003	-0.0006*
	(0.0003)	(0.0003)
SIZE	2.1297***	2.1939***
	(0.2406)	(0.2802)
ROA	31.0117***	24.3053***
	(2.7138)	(4.2397)
LOSS	-1.5826***	-2.8526***
	(0.4870)	(0.5189)
LEV	-8.3066***	-8.8792***
	(1.3966)	(2.0336)
INT_COV	0.0176	0.0187
	(0.0123)	(0.0116)
CAP_INTEN	-0.7414	0.0535
	(0.6250)	(0.8268)
BIG4	2.3424***	-0.6315
	(0.8370)	(0.6658)
CFO_STD	-12.2237*	-39.6285***
	(7.2357)	(10.6845)
ROA_STD	-16.2214***	-26.7302***
	(3.3983)	(6.6894)
MARGIN	2.8571*	11.5162***
	(1.6985)	(2.2872)
Time F.E.	YES	YES
Agency F.E.	YES	YES
Industry F.E.	YES	YES
Country F.E.	NO	YES
Firm Clustered	YES	YES
Observations	243,259	104,671
Adjusted R^2	0.484	0.513

Notes: this table reports the results from OLS regression of credit ratings on the environmental score and the square of environmental score. We use the numerical transformation of foreign long-term issuer ratings by S&P, Moody's, and Fitch, increasing in credit quality. $RATING\_58$ is the credit rating scaled from 0 to 58, while ENV is the environmental score provided by Thomson Reuters ASSET4. All regressions include year-month, agency, and industry fixed effects, while country fixed effects are only employed in the EU sample. Reported significance is based on robust standard errors clustered at firm level. Significance at 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. † indicates that, performing an one-sided significance test, the parameter estimate is significantly larger than zero at 10% significance level.

the minimum environmental score) to 0 (evaluated at 91.25) and it even becomes negative.

Diverse regulatory environments and market perceptions in the EU and US may explain the detected discrepancies in the link between environmental scores and credit ratings. In the US, firms cluster at lower environmental scores, hence those achieving high environmental performance are often viewed as pioneering protectors of the environment, resulting in a more noticeable positive impact on their credit ratings. Conversely, in the EU, where environmental regulations are stricter and a larger proportion of firms attain high environmental scores, being environmentally conscious might be seen as a baseline expectation, rather than a distinguishing factor. This sheds light on the left-skewed distribution of environmental scores and diminishing marginal effect observed in the EU as environmental scores increase: firms are still rewarded for improved environmental performance, but the magnitude of the reward diminishes.[11]

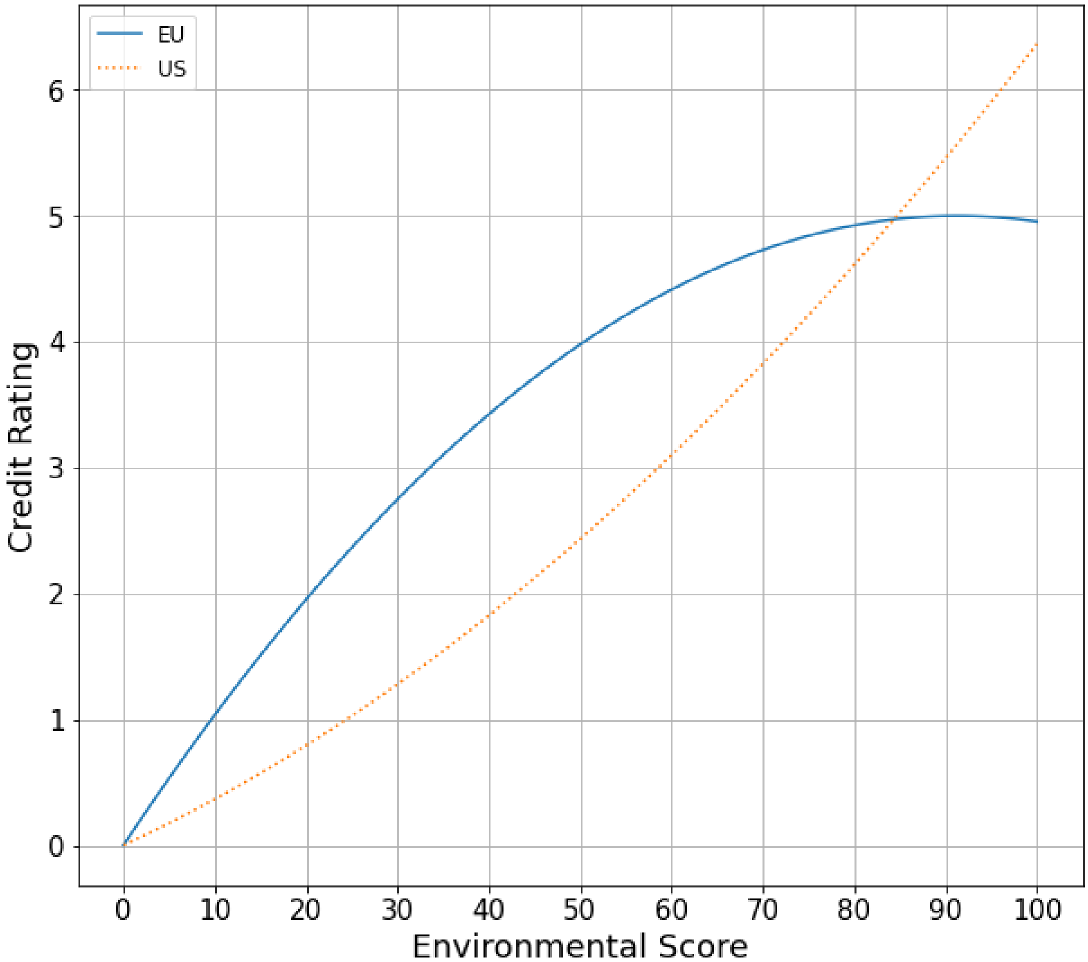
4.6 Endogeneity tests

In this section we present the results of tests to address the potential endogeneity issues. Additionally, we employ instrumental variable estimation method to address the endogeneity concern.

4.6.1 Test for omitted variable bias

One concern with our analysis is that relevant variables might have been omitted from our model. To assess whether omitted variable bias is present, we carry out

Figure 4.5.3: Environmental performance versus credit rating



Note: This figure presents the non-linear relationship between credit ratings and environmental score. The horizontal axis represents the environmental score. The vertical axis represents the predicted value of the numerical transform of credit rating scaled from 0 to 58. The solid blue (dotted orange) line shows the relationship in the EU (US). The figure is based on the parameter estimates of Eq. (4.5.1) reported in Table 4.5.6. For simplicity, the control variables are held at zero.

a test proposed by Oster (2019), for our OLS regression results displayed in Table 4.5.4, Panel A, Columns (1) and (2). This test addresses the stability of regression coefficients and R-squared with and without controls to establish an identifiable set for the coefficient of interest. If zero is not included in this set, then the null hypothesis that an omitted variable is driving the result can be dismissed. One boundary of this identifiable set is $\tilde{\beta}$, the coefficient of interest in the model with controls. The other bound, denoted as β^*, is computed as follows:

$$\beta^* \approx \tilde{\beta} - \delta[\dot{\beta} - \tilde{\beta}]\frac{R_{max} - \tilde{R}}{\tilde{R} - \dot{R}}, \tag{4.6.1}$$

where $\dot{\beta}$ stands for the coefficient of interest from the regression without control variables. $\tilde{R}$ is the R^2 when all controls are included while $\dot{R}$ is the R^2 without controls. Oster (2019) suggests that a suitable upper bound for δ is 1, although no standard approach exists. R_{max} symbolizes the R-squared of a hypothetical model including both observable and unobservable covariates, and is suggested to be $R_{max} = 1$ (the most stringent case), $R_{max} = min(2\tilde{R}; 1)$, or $R_{max} = min(1.5\tilde{R}; 1)$. Using the coefficient of environmental score from US and EU samples in Table 4.5.4, Panel A Columns (1) and (2) as the upper bounds, as well as the coefficient of environmental score and R^2 without any control variables as the lower bounds, we construct Oster's identifiable set, with $\dot{\beta}$ for US (EU) samples being 0.1230 (0.0912) and $\tilde{R}$ for US (EU) samples being 0.1302 (0.0645). Assuming $\delta = 1$ and $R_{max} = 1$, the identifiable set for the EU sample regression is $[0.0049, 0.0498]$, excluding zero in the most stringent case. Assuming $\delta = 1$ and $R_{max} = min(1.5\tilde{R}; 1)$, the identifiable set for the US sample regression is $[0.0118, 0.0570]$. Oster comments that employing $\delta = 1$ and $R_{max} = 1$ results in only about one-third of empirical research studies in leading economic journals being robust, thus suggesting less restrictive alternatives such as $R_{max} = min(1.5\tilde{R}; 1)$ as acceptable.

4.6.2 Instrumental variable estimation

In our second endogeneity test, we verify the stability of our evidence to potential endogeneity bias stemming from reverse causality. One might argue that firms with better credit ratings can support more environmental-related investments. We control for this potential bias by employing the two-stage least squares (2SLS) regression to examine whether our results are driven by endogeneity between environmental scores and credit ratings. In the analysis, we use an instrumental variable labelled as IV_INDUS, which represents the average monthly environmental score of the firms in a given industry industry. This is directly related to firm-level environmental scores within the industry but it holds no direct connections to the individual credit ratings.

The 2SLS regression results are presented in Table 4.6.7, showing the first and second stage regression results for both the US and EU samples. The first stage of the 2SLS regression indicates a significant positive relation between the instrumental variable and firm-level environmental score, suggesting that the instrumental variable is valid for the study. In the second stage, the instrumented environmental score is used in the regression analysis with credit ratings. The results show a positive and significant relation between the instrumented environmental score and credit ratings, with the coefficient being significant at 1%

Table 4.6.7: Instrumental variable (2SLS) results

	(1)	(2)	(3)	(4)
	US		EU	
	First-stage	Second-stage	First-stage	Second-stage
Variables	Second-stage Dependent Variable = RATING_58			
ENV		0.0603*		0.1327***
		(0.0355)		(0.0450)
IV_INDUS	0.6173***		0.6606***	
	(0.0751)		(0.0738)	
SIZE	11.9570***	2.0971***	8.9396***	1.4074***
	(0.4381)	(0.4708)	(0.6114)	(0.5043)
ROA	59.1454***	30.8919***	33.5890***	21.1324***
	(8.0570)	(3.5440)	(10.6274)	(4.4218)
LOSS	2.1582	-1.5767***	1.4704	-2.9701***
	(1.3735)	(0.4900)	(1.4231)	(0.5298)
LEV	-8.3104**	-8.1872***	3.5638	-9.2850***
	(3.7701)	(1.4126)	(5.0859)	(2.0413)
INT_COV	0.0102	0.0178	0.0268	0.0168
	(0.0236)	(0.0123)	(0.0288)	(0.0127)
CAP_INTEN	10.1156***	-0.7868	4.6577**	-0.3382
	(2.0250)	(0.7195)	(2.0792)	(0.8795)
BIG4	6.4841***	2.2390***	2.2123	-0.8933
	(2.2284)	(0.8674)	(2.1225)	(0.7478)
CFO_STD	-16.7461	-11.7075	-16.2546	-39.0591***
	(21.5907)	(7.3427)	(29.3178)	(10.6166)
ROA_STD	27.4377***	-16.4053***	-9.9813	-25.3562***
	(10.2650)	(3.5738)	(12.8412)	(6.4896)
MARGIN	-22.7740***	2.9481	-10.6332*	12.6400***
	(4.8315)	(1.9056)	(5.7656)	(2.3972)
CONSTANT	-118.9133***	5.9092*	-65.8254***	13.7237***
	(6.2754)	(3.2309)	(9.3967)	(3.4955)
Time F.E.	YES	YES	YES	YES
Agency F.E.	YES	YES	YES	YES
Industry F.E.	YES	YES	YES	YES
Country F.E.	NO	NO	YES	YES
Firm Clustered	YES	YES	YES	YES
Observations	243,259	243,259	104,671	104,671
Underidentification test				
Kleibergen-Paaprk LM statistic		56.33***		46.64***
Weak identification test				
Cragg-Donald Wald F statistic		3342.230***		2683.453***
Weak-instrument-robust inference				
Anderson-Rubin Wald test		3.11*		8.33***

Notes: This table reports the results of two-stage least square regressions for US and EU samples in Columns (1)–(2) and (3)–(4), respectively. We use the numerical transformation of foreign long-term issuer ratings by S&P, Moody's, and Fitch, increasing in credit quality. $RATING\_58$ is the credit rating scaled from 0 to 58, details in Section 4.3.1. ENV is the environmental score provided by Thomson Reuters ASSET4. IV_INDUS is the monthly average of the environmental score of firms in a given industry. All regressions include year-month, agency, and industry fixed effects, while country fixed effects are only employed in the EU sample. Reported significance is based on robust standard errors clustered at firm level. Significance at 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

level in the EU. For the US sample, it is significant at 10% level. We also conduct several tests to further validate the use of the instrumental variable. The results of the underidentification test (Kleibergen-Paaprk LM-statistic) and the weak identification test (Cragg-Donald Wald F-statistic) show that the instrumental variable is strong and relevant. The Anderson-Rubin Wald test further confirms that the instrumented variable is robust to the weak instrument bias. All these tests support the choice of the instrumental variable and the validity of the results of the study. Thus, the analysis supports the hypothesis that better environmental performance, as measured by environmental scores, is positively associated with credit ratings.

4.6.3 Propensity score matching & entropy balancing

The existing literature documents that firms with better financial performance have higher ESG/CSR performance (Hong et al., 2012; Borghesi et al., 2014). One may argue that firms with greater financial performance have the ability to expend more resources on ESG/CSR activities. This means that the ESG scores of financially well-performing companies could differ from those which are underperforming. To address the potential differences between firms with high and low environmental performance, we employ the propensity score matching model (PSM) developed by Rosenbaum and Rubin (1983) to address the concern that the treated sample is not similar to the control (see Fang et al., 2014). Unlike conventional selection models such as the one proposed by Heckman (1979) that estimate the effects of treatments based on certain functions, PSM does not make any assumptions about functional relation. Instead, it provides a more direct way to estimate the effect of treatments (see Kai and Prabhala, 2007).

To implement this approach in our study, we first divide our sample into two subsamples based on the median environmental score. Firms scoring above (below) the median are defined as the treatment (control) group. Similar to Table 4.5.5, $HIGH\_ENV$ is used. This variable equals one when the firm is part of the treatment group and zero otherwise. A logit model is used to estimate the propensity score using control variables from the baseline regression, agency indicators, and year-month indicators: $SIZE$, ROA, $LOSS$, LEV, INT_COV, CAP_INTEN, $BIG4$, CFO_STD, ROA_STD, $MARGIN$, *Agency dummies*, and *Time dummies*. Ultimately, we perform a one-to-one matching, allowing a maximum caliper distance of 1% without replacement (Lawrence et al., 2011; Shipman et al., 2017).

Table 4.6.8 (Panel A) presents the OLS and ordinal logit regression results using samples treated with PSM. When comparing observations in PSM with our primary results in Panel A of Table 4.5.4, nearly half of the observations are eliminated after the matching, and this removal rate is similar for both the US and EU samples. Nonetheless, our findings consistently show a positive and significant coefficient for both OLS and ordinal logit regressions in the US and EU samples. These results are consistent with our main regression results in Table 4.5.4 (Panel A) that firms with higher environmental score exhibit higher credit ratings.

Although PSM offers an effective approach to address endogeneity concerns, one criticism of PSM is that variations in design choices, including maximum caliper width, matching with/without common support, with/without replace-

Table 4.6.8: Test for propensity score matching and entropy balancing

	(1)	(2)	(3)	(4)
	US	EU	US	EU
	PSM		Entropy Balancing	
Variables	Dependent Variable = RATING_58			
ENV	0.0523***	0.0397***	0.1400***	0.1243***
	(0.0084)	(0.0146)	(0.0096)	(0.0147)
SIZE	2.1876***	2.0636***	-0.0565	0.3544***
	(0.2351)	(0.3038)	(0.1425)	(0.0822)
ROA	31.1018***	25.2797***	10.5263***	18.3535***
	(3.1000)	(5.1699)	(2.7406)	(4.2356)
LOSS	-1.7555***	-2.9791***	-2.4628***	-2.3367***
	(0.5554)	(0.5853)	(0.8476)	(0.6163)
LEV	-10.1247***	-11.7249***	-4.5664***	-8.2097***
	(1.3794)	(2.3521)	(1.2494)	(1.8527)
INT_COV	0.0204*	0.0153	-0.0031	0.0140
	(0.0109)	(0.0144)	(0.0073)	(0.0127)
CAP_INTEN	-0.9717	0.2836	-0.5430	-0.6709
	(0.6932)	(0.8504)	(0.4748)	(0.6291)
BIG4	3.9811***	-0.6117	2.0509	-1.0008*
	(1.2078)	(0.6931)	(1.7265)	(0.5672)
CFO_STD	-17.6050**	-20.5286*	2.8375	-20.8999**
	(7.5569)	(12.3927)	(7.6702)	(9.1136)
ROA_STD	-16.6919***	-35.3795***	-20.1634***	-36.8475***
	(4.0673)	(8.5785)	(4.3772)	(7.8634)
MARGIN	4.4787***	12.0939***	3.2416***	11.4335***
	(1.6583)	(2.2142)	(1.1045)	(2.1802)
Time F.E.	YES	YES	YES	YES
Agency F.E.	YES	YES	YES	YES
Industry F.E.	YES	YES	YES	YES
Country F.E.	NO	YES	NO	YES
Firm clustered	YES	YES	YES	YES
Observations	131,942	64,122	243,259	104,671
Adj. R^2/Pseudo R^2	0.411	0.450	0.362	0.406

Notes: This table reports the OLS regression results of the sample constructed using propensity score matching (PSM) and entropy balancing. We use the numerical transformation of foreign long-term issuer ratings by S&P, Moody's, and Fitch, increasing in credit quality. $RATING\_58$ is the credit rating scaled from 0 to 58, details in Section 4.3.1. ENV is the environmental score provided by Thomson Reuters ASSET4. All regressions include year-month, agency, and industry fixed effects, while country fixed effects are only employed in the EU sample. Reported significance is based on robust standard errors clustered at firm level. Significance at 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

ment, and whether one-to-one matching is used or not can influence the conclusions (see DeFond et al., 2016). Another criticism is that unmatched units from PSM are discarded, reducing the number of observations for subsequent tests (Wilde, 2017; Chapman et al., 2019).

For these reasons, we also implement entropy balancing, an alternative technique to tackle endogeneity concerns.[12] This method, which does not require a model specification or criteria, is a weighting technique designed to improve balance between the treatment and control groups without losing observations (Hainmueller, 2012; Hainmueller and Xu, 2013). Specifically, this method adjusts the weights of the control group observations such that the first, second, and third moment (i.e., mean, standard deviation, and skewness) of all covariates in the control group to match those of the treatment group. As shown in Panel B of Table 4.6.8, we continue to observe a significant and positive relation between credit ratings and environmental scores. Compared to the results in Panel A, the regression results from a complete sample indicate that the magnitudes of all coefficients of our variable of interest (ENV) are larger than those from the PSM-matched sample.

4.7 Robustness tests

We carry out several additional studies to prove the robustness of our results, including re-running regressions by distinguishing between investment-grade and speculative-grade ratings, re-running regressions for individual rating agencies, replacing foreign-currency issuer ratings with domestic-currency ratings, employing alternative measures for the environmental performance, industry-size matched sampling, as well as investigating the non-linear relation of social and governance performance on credit ratings.

4.7.1 Investment Grade versus Speculative Grade

Following Ashbaugh-Skaife et al. (2006), Cornaggia et al. (2017), and Bhandari and Golden (2021), we employ an alternative measure for credit ratings. We construct a dummy variable (INVESTMENT_GRADE), which is assigned 1 if the long-term issuer credit rating falls in the top tier (BBB- or above), and 0 otherwise. We apply the logit model for our regression analysis. We aim to evaluate whether firms with above median-level environmental scores have a higher likelihood of receiving an investment-grade rating compared to those with lower environmental scores. We also incorporate the binary variable $HIGH\_ENV$ in this analysis. The logit regression results are reported in Table 4.7.9. The coefficients on ENV are positive and statistically significant, with a value of 0.0212 (0.0144) for the US (EU) sample. However, the coefficient of $HIGH\_ENV$ is 0.7839 and significant at 1% level for the US, while in the EU it is only 0.3220 and significant at 10% level only. This suggests that EU firms with higher or lower environmental scores do not show as significant differences as US firms, aligning with our Hypothesis 2.

4.7.2 Regression analysis by CRA

As credit ratings might vary across rating agencies due to differing rating methodologies, we study the effect of environmental scores on credit ratings by running

Table 4.7.9: Results for investment grade dummy

	(1)	(2)	(3)	(4)
	US		EU	
Variables	Dependent Variable = INVESTMENT_GRADE			
ENV	0.0212***		0.0144***	
	(0.0033)		(0.0053)	
HIGH_ENV		0.7839***		0.3220*
		(0.1290)		(0.1823)
Controls	YES	YES	YES	YES
Time F.E.	YES	YES	YES	YES
Agency F.E.	YES	YES	YES	YES
Industry F.E.	YES	YES	YES	YES
Country F.E.	NO	NO	YES	YES
Firm Clustered	YES	YES	YES	YES
Observations	243,259	243,259	104,340	104,340
Adj. R^2/Pseudo R^2	0.385	0.376	0.356	0.351

Notes: This table reports the coefficients of logit regression for the US and EU samples. The dependent variable is *INVESTMENT_GRADE*, an indicator variable which equals one if the long-term issuer credit rating is in the top group (also known as investment-grade BBB- or higher), and zero otherwise (the bottom group, also known as the speculative-grade BB+ or lower). We use the long-term foreign currency issuer ratings by S&P, Moody's, and Fitch. *ENV* is the environmental score provided by Thomson Reuters ASSET4. *HIGH_ENV* equals one if the environmental score is above the median level of the environmental score, zero otherwise. All regressions include year-month, agency, and industry fixed effects, while country fixed effects are only employed in the EU sample. Reported significance is based on robust standard errors clustered at firm level. Significance at 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

CRA-specific regressions. The results are reported in Table 10. We find solid evidence that environmental scores positively impact credit ratings across all subsets, validating our Hypothesis 1 that higher environmental performance leads to a higher credit rating. However, the strength of this relation varies across different rating agencies and regions. For instance, Moody's displays the highest environmental coefficient in the US with a value of 0.072, compared to 0.046 in the EU, while Fitch exhibits a higher coefficient in the EU than in the US. The results are consistent across both OLS regressions (Panel A) and ordinal logit regressions (Panel B).[13]

4.7.3 Domestic-currency ratings

We analyze whether the positive relation between environmental scores and ratings changes if we use domestic-currency issuer credit ratings as the dependent variable instead of foreign-currency ratings. The rationale is that foreign currency ratings could incorporate exchange rate and inflation risks, which are absent in domestic currency ratings, and could potentially weaken the correlation between environmental scores and foreign currency ratings. Table 4.7.11 presents the regression results using domestic currency credit ratings as the dependent variable. Compared to the baseline regression results in Table 4.5.4, the coefficients of all regressions with domestic currency ratings are higher than those with foreign currency ratings. As expected, excluding potential risks from exchange rates and inflation yields a higher coefficient, with the difference being more pronounced when employing OLS regressions with 58 scaled ratings in columns 1 and 2.

Table 4.7.10: Results for rating subsamples according to credit rating agencies

Panel A: OLS regressions						
	(1)	(2)	(3)	(4)	(5)	(6)
	US			EU		
	SP	Moody's	Fitch	SP	Moody's	Fitch
Variables	Dependent Variable = RATING_58					
ENV	0.0566***	0.0720***	0.0304**	0.0555***	0.0460**	0.0412**
	(0.0092)	(0.0117)	(0.0152)	(0.0171)	(0.0201)	(0.0192)
Controls	YES	YES	YES	YES	YES	YES
Time F.E.	YES	YES	YES	YES	YES	YES
Industry F.E.	YES	YES	YES	YES	YES	YES
Country F.E.	NO	NO	NO	YES	YES	YES
Firm Clustered	YES	YES	YES	YES	YES	YES
Observations	111877	77273	54109	48060	30972	25638
Adj. R^2/Pseudo R^2	0.495	0.443	0.497	0.531	0.514	0.543

Panel B: Ordinal logit regressions						
	(1)	(2)	(3)	(4)	(5)	(6)
	US			EU		
	SP	Moody's	Fitch	SP	Moody's	Fitch
Variables	Dependent Variable = RATING_20					
ENV	0.0162***	0.0181***	0.0113***	0.0125***	0.0113***	0.0129***
	(0.0021)	(0.0025)	(0.0032)	(0.0041)	(0.0055)	(0.0064)
Controls	YES	YES	YES	YES	YES	YES
Time F.E.	YES	YES	YES	YES	YES	YES
Industry F.E.	YES	YES	YES	YES	YES	YES
Country F.E.	NO	NO	NO	YES	YES	YES
Firm Clustered	YES	YES	YES	YES	YES	YES
Observations	111,877	77,273	54,109	48,060	30,973	25,638
Adj. R^2/Pseudo R^2	0.174	0.141	0.160	0.188	0.155	0.188

Notes: This table reports the results from regressions of credit ratings on environmental score for subsamples of different credit rating agencies. We use the numerical transformation of foreign long-term issuer ratings by S&P, Moody's, and Fitch, increasing in credit quality. $RATING\_58$ is credit rating scaled from 0 to 58, while $RATING\_20$ is scaled from 1 to 20, details in Section 4.3.1. ENV is the environmental score provided by Thomson Reuters ASSET4 ESG database. Panel A reports coefficients estimated using OLS regressions with credit ratings that are scaled from 0 to 58, while Panel B reports the ordinal logit regression results with credit ratings scaled from 1 to 20. All regressions include year-month, agency, and industry fixed effects, while country fixed effects are only employed in the EU sample. Reported significance is based on robust standard errors clustered at firm level. Significance at 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Table 4.7.11: Results for credit ratings with domestic currency

	(1)	(2)	(3)	(4)
	US	EU	US	EU
	OLS		Ordinal Logit	
Variables	Dependent Variable = Domestic RATING_58		Dependent Variable = Domestic RATING_20	
ENV	0.0593***	0.0505***	0.0159***	0.0118***
	(0.0084)	(0.0143)	(0.0021)	(0.0039)
Controls	YES	YES	YES	YES
Time F.E.	YES	YES	YES	YES
Agency F.E.	YES	YES	YES	YES
Industry F.E.	YES	YES	YES	YES
Country F.E.	NO	YES	NO	YES
Firm Clustered	YES	YES	YES	YES
Observations	244,134	104,214	244,134	104,214
Adj. R^2/Pseudo R^2	0.486	0.511	0.157	0.169

Notes: This table reports the results from regressions of credit ratings on environmental score. We use the numerical transformation of domestic long-term issuer ratings by S&P, Moody's, and Fitch, increasing in credit quality. $RATING\_58$ is rating scaled from 0 to 58, while $RATING\_20$ is scaled from 1 to 20, details in Section 4.3.1. ENV is the environmental score provided by Thomson Reuters ASSET4 ESG database. All regressions include year-month, agency, and industry fixed effects, while country fixed effects are only employed in the EU sample. Reported significance is based on robust standard errors clustered at firm level. Significance at 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

4.7.4 Alternative measures for environmental performance

To corroborate the validity of our primary results, we perform a robustness check by employing alternative measures for the environmental scores. The necessity of such robustness check is to ensure that our findings are not solely dependent on the particular measure of environmental score used. For our analysis, we turn to two key alternatives: Green House Gas (GHG) emissions and Bloomberg's environmental scores.

GHG emissions, as identified in the Thomson Reuters ASSET4 ESG database, constitute a major determinant of the overall environmental score. We consider the logarithm of total CO_2 and CO_2 equivalent emissions in tonnes, which includes both direct and indirect emissions from owned or controlled sources, as our first alternative measure for environmental performance. We re-run the baseline regression using this variable (denoted $EMISSION$) and the results are presented in Table 4.7.12. Given the implied inverse relation between emissions and environmental friendliness, it is expected that the coefficients for $EMISSION$ are negative and statistically significant. In line with this expectation, the results from the EU sample are consistently more negative than those from the US. This suggests that credit rating agencies are incorporating information on carbon emissions when evaluating credit ratings, and this is more pronounced in the EU than in the US.

Next, we use the environmental scores issued by Bloomberg as our second alternative measure for environmental performance. Since Bloomberg's environmental scores range from 0 to 10, unlike the 0-100 scale used in Thomson Reuters ASSET4, the coefficients obtained based on the scores issued by Bloomberg are typically larger than those from our baseline regressions.

4.7.5 Industry and size matched sampling

As opposed to the PSM methodology used in section 4.6.3 and motivated by Bhandari and Golden (2021), here we specifically focus on two variables, industry and size, to get a better understanding on their effect on the results. Since the distribution of environmental scores might vary across industries, we employ an industry- and size-matched sample with the high environmental scores group and the low environmental scores group to fully capture differences across industries. By using the $HIGH\_ENV$ dummy variable introduced at the beginning of Section 4.5, we obtain the same effect as dividing our sample into two subsam-

Table 4.7.12: Results using alternative environmental performance proxies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	US		EU		US		EU	
	OLS				Ordinal Logit			
Variables	Dependent Variable = RATING_58				Dependent Variable = RATING_20			
EMISSION	-0.6259**		-0.6708***		-0.1436**		-0.1946**	
	(0.2487)		(0.2378)		(0.0596)		(0.0803)	
ENV_BlOOMBERG		0.4319***		0.2849*		0.1318***		0.1119**
		(0.1400)		(0.1566)		(0.0352)		(0.0545)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Time F.E.	YES	YES	YES	YES	YES	YES	YES	YES
Agency F.E.	YES	YES	YES	YES	YES	YES	YES	YES
Industry F.E.	YES	YES	YES	YES	YES	YES	YES	YES
Country F.E.	NO	NO	YES	YES	NO	NO	YES	YES
Firm Clustered	YES	YES	YES	YES	YES	YES	YES	YES
Observations	141,743	190,070	91,732	72,569	141,743	190,070	91,733	72,569
Adj. R^2/Pseudo R^2	0.424	0.464	0.524	0.507	0.140	0.150	0.179	0.182

Notes: This table reports the results from regressions of credit ratings on various environmental measures. We use the numerical transformation of foreign long-term issuer ratings by S&P, Moody's, and Fitch, increasing in credit quality. $RATING\_58$ is the credit rating scaled from 0 to 58, while $RATING\_20$ is scaled from 1 to 20, details in Section 4.3.1. $EMISSION$ is the logarithm of the total CO_2 emission (the CO_2 emission scope 1 plus scope 2) provided by Thomson Reuter ASSET 4, while $ENV\_BLOOMBERG$ represents the environmental score provided by Bloomberg ranging from 0 to 10. All regressions include year-month, agency, and industry fixed effects, while country fixed effects are only employed in the EU sample. Reported significance is based on robust standard errors clustered at firm level. Significance at 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

ples based on the median level of ENV. Re-running the baseline regression with industry and size matched sampling, the results are presented in Table 4.7.13. The implications from this analysis suggest that our results are not sensitive to industry variations.

Table 4.7.13: Results based on industry and size matched sampling

	(1)	(2)	(3)	(4)
	US		EU	
Variables	Dependent Variable = RATING_58			
ENV	0.0524***		0.0445***	
	(0.0086)		(0.0139)	
HIGH_ENV		1.9374***		1.0759***
		(0.3405)		(0.4014)
Controls	YES	YES	YES	YES
Time F.E.	YES	YES	YES	YES
Agency F.E.	YES	YES	YES	YES
Industry F.E.	YES	YES	YES	YES
Country F.E.	NO	NO	YES	YES
Firm Clustered	YES	YES	YES	YES
Observations	93,884	93,884	38,966	38,966
Adj. R^2	0.439	0.433	0.48	0.476

Notes: This table reports the OLS regression results of the sample constructed using industry-size Matching. We use the numerical transformation of foreign long-term issuer ratings by S&P, Moody's, and Fitch, increasing in credit quality. $RATING\_58$ is the credit rating scaled from 0 to 58, details in Section 4.3.1. ENV is the environmental score provided by Thomson Reuters ASSET4. $HIGH\_ENV$ is an indicator variable which equals one if the environmental score is above the median, otherwise 0. All regressions include year-month, agency, and industry fixed effects, while country fixed effects are only employed in the EU sample. Reported significance is based on robust standard errors clustered at firm level. Significance at 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

4.7.6 Effects from social and governance performance

To further investigate whether the non-linear relation in Table 4.5.6 only exists between firms' environmental performance and credit ratings, we replace the environmental score (ENV) with the social ($SOCIAL$) and governance (GOV) scores provided by the Thomson Reuter ASSET4 ESG ratings. Re-running the nonlinear regressions from the beginning of section 5, the regression results are presented in Table 4.7.14. Our findings, which align with our expectations, show no statistical evidence of a non-linear relation between the other ESG components ($SOCIAL$ and GOV) and credit ratings. These results confirm the uniqueness of the relation between environmental performance and credit ratings, thereby emphasizing that it is not a common attribute of ESG performance.

4.8 Conclusion

Credit rating agencies play an essential role in the financial markets by, issuing assessments of the companies' creditworthiness. In recent years, CRAs started to include environmental aspects into their rating assessment due to the increasing importance of firms' environmental performance. Inspired by the increasing global awareness of environmental sustainability, our study introduces a transatlantic perspective by investigating the impact of the firms' environmental performance on their credit ratings in the US and the EU. Considering differentiated regulatory requirements of ESG/CSR between the two economies, we further examine whether the influence of environmental performance on credit ratings differs across these two regions.

Table 4.7.14: Results based on social and governance scores

	(1)	(2)	(3)	(4)
	US		EU	
Variables	Dependent Variable = RATING_58			
SOCIAL	0.0377		0.0251	
	(0.0344)		(0.0501)	
GOV		0.0563**		-0.0421
		(0.0281)		(0.0348)
SOCIAL^2	0.0003		0.0001	
	(0.0004)		(0.0004)	
GOV2^2		-0.0002		0.0004
		(0.0003)		(0.0003)
Controls	YES	YES	YES	YES
Time F.E.	YES	YES	YES	YES
Agency F.E.	YES	YES	YES	YES
Industry F.E.	YES	YES	YES	YES
Country F.E.	NO	NO	YES	YES
Firm clustered	YES	YES	YES	YES
Observations	243,259	243,259	104,671	104,671
Adj. R^2	0.485	0.472	0.507	0.503

Notes: This table reports the OLS regression results on the effects on credit ratings of the social and governance score and their quadratic terms. We use the numerical transformation of foreign long-term issuer ratings by S&P, Moody's, and Fitch, increasing in credit quality. $RATING\_58$ is the credit rating scaled from 0 to 58, while $SOCIAL$ and GOV are the social and governance scores provided by Thomson Reuters ASSET4 and $SOCIAL^2$ and GOV^2 are the squared social and governance scores. All regressions include year-month, agency, and industry fixed effects, while country fixed effects are only employed in the EU sample. Reported significance is based on robust standard errors clustered at firm level. Significance at 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Our baseline analysis explores the effect of environmental performance on credit ratings. Our analysis uses numerical ratings that account for the rating outlook and watch. Our findings suggest that an improvement of firms' environmental scores contributes to higher credit ratings. However, we note a weaker relation in the EU compared to the US. We undertake additional investigation to corroborate our initial analysis. Our results indicate the main cause for this weaker effect: the effect of environmental performance on credit ratings is non-linear in the EU, resulting in the diminishing marginal effect of environmental score improvement on credit ratings. This is because firms in the EU being environmentally friendly might be viewed more as a norm, rather than a stand-out performance. Thus, improvements in environmental scores are less rewarded (in terms of credit rating improvements) for EU firms with good environmental performance.

Our empirical results have significant implications for corporate financial management. Besides the profitability-related factors that can improve a firm's credit rating, we reveal an additional way on which firms can enhance their creditworthiness by improving their environmental performance. Therefore, firms can reduce financing costs via the channel of better environmental performance. Our result also suggest that this channel to reduce financing costs is more effective in the US than in the EU.

Our study also opens up avenues for further exploration. Some of the potential extensions of our research include: study of the relation between environmental scores and credit ratings across different industries; analysis of the dynamic of this relation across time, as well as assessment of the influence of social and governance factors on credit ratings. We see these questions as opportunities to enrich the literature and broaden our understanding, and we leave these to be explored in future research.

Notes

[1]We follow Gillan et al. (2021) to treat the terms ESG and CSR as if they are interchangeable and use the terminology ESG/CSR.

[2]https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets_en, assessed on 03 July 2023.

[3]The final EU sample incorporates data from 20 EU countries, including the United Kingdom, France, Germany, Italy, Netherlands, Spain, Sweden, Finland, Luxembourg, Ireland, Portugal, Austria, Denmark, Belgium, Greece, Czech Republic, Hungary, Cyprus, Romania, and Slovenia.

[4]Although $RATING\_20$ is a categorical variable, we keep it in Table 4.3.2 for statistical purposes.

[5]Variables used in the multicollinearity test are $RATING\_58$, ENV, and all firm-level control variables.

[6]The two numerical transformations of credit rating are employed with different research purposes, $RATING\_58$ is treated as a pseudo-continuous variable for linear regression, while $RATING\_20$ is treated as categorical variable for ordinal logit regression. Thus, no conclusion on which one of the numerical transformations is better.

[7]The Big4 auditor are the four largest global accounting networks in the world: Deloitte, Ernst & Young (EY), KPMG, and PwC.

[8]The industry and country classification in this study is based on the Industry Classification Benchmark (ICB) and ISO country code, respectively.

[9]We are comparing the coefficients obtained from the US and EU data, but we acknowledge that there are distributional differences so the interpretation of the differences is not straightforward.

[10]An environmental score of 0 represents poor environmental performance. Missing values are marked as NaN, which are removed before statistical analysis.

[11]Lahouel et al. (2022) also find an inverted U-shaped relationship between environmental performance and financial performance ranging in France, and non-linear relationships in other European countries.

[12]Another alternative methodology that can be considered is based on Bartram et al. (2022) and we leave this for future research.

[13]The diverse results from different credit rating agencies also reduce the possibility that a change in rating by one agency tends to be followed by changes by the other agencies.

Chapter 5

Conclusions and Further Research

5.1 Summary of the Findings

This thesis delivers critical insights into risk measurement that can greatly benefit risk managers, regulators, investors, and other industry professionals.

In Chapter 2 we introduce an innovative set of market risk models, which are an extension of the dynamic semiparametric models proposed by Patton et al. (2019) designed to jointly forecast two crucial risk measures: VaR and ES. These models incorporate information from an auxiliary level of significance to enhance the estimation of risk at extreme percentiles. One of the key techniques employed in the A-GAS models is TSCV, which is used to select the optimal auxiliary level of significance. The results show that TSCV is highly effective for this purpose, leading to a substantial improvement in the forecasting performance of risk measures as highlighted by a simulation study and an empirical analysis.

In the simulation study, the A-GAS models consistently outperform the original GAS models. Particularly, the A-GAS models achieved lower backtest rejection rates and loss values, signifying more accurate risk forecasts. In estimating the in-sample parameters, we show that the VaR and ES produced by the A-GAS models can generate relatively lower in-sample losses compared with those from original models. The out-of-sample results also show that the A-GAS models consistently outperform the original GAS models and other prevailing benchmarks across various backtests based on data on different oil futures. We observe an even more pronounced superior performance during COVID-19 period.

In Chapter 3, we investigate the impact from the climate transition risk to total downside risk of equity (measured by VaR and ES). We proxy for the climate transition risk factors by employing the environmental scores from the Thompson Reuter ASSET 4 ESG database, namely Emission score, Innovation score, and Resource Use score. We first examine the relationship between transition climate risk factors and stock returns using panel quantile regressions. Our empirical results show a negative relationship between them in the low quantiles of the stock returns, which implies that financially underperforming companies are negatively affected by the cost of improving their environmental scores. To further investigate the effects of environmental scores on downside risk, we focus our analysis on the loss of companies by regressing VaR and ES at 1% level on the climate risk factors for 11 sectors. Our main finding is that corporate investments in improvements of their environmental scores reduce their total downside risk in the Energy and Utilities sectors, whilst it increases their total downside risk in the Health Care sector. Based on these regression results, we propose a novel set of risk measures (climate VaR and climate ES) that capture the market risk attributed to climate transition risk factors, and find heterogeneity in the sensitivity of the firm-level risk to environmental scores across sectors. These findings have important implications for investors and business managers who are concerned about the impact from climate risks on their financial portfolios.

In Chapter 4, we examine whether improvements in corporate environmental performance has a positive impact on the firms' credit ratings. In particular, our study conduct a transatlantic study covering companies in the US and EU to explore any differences in the nature of this relationship between the two regions. Data of corporate environmental performance (proxied by the environmental score) is sourced from the Thomson Reuters ASSET4 ESG database, while the credit rating sample is extracted from Bloomberg. We not only convert credit ratings to a 20-point numerical scale, but also extend it to a 58-point scale system by combining the outlook and watch signal with rating changes. We employ OLS estimation (ordinal logit model) with the 58- (20-) point scaled credit ratings as the dependent variable by controlling for several firm characteristics. We find a significantly positive relationship between corporate environmental performance and credit ratings in both regions, with a higher environmental performance and greater marginal effects across different ratings in the US compared to the EU. To further investigate the regional difference, we examine potential nonlinear relationships between environmental scores and credit ratings by adding a quadratic term of environmental scores. In this setup, we find that in the EU the relationship of the environmental score and credit ratings is concave downwards. Additionally, through rigorous endogeneity and robustness tests using diverse methodologies, our primary results have been consistently validated. Overall, our findings shed light on the implications of environmental performance and provide critical insights for firms seeking to improve their credit rating via sustainability initiatives.

5.2 Suggestions for Future Research

While this thesis makes a significant contribution to the measurement of financial risks, such as market risk, climate risk, and credit risk, there remains ample room for further exploration and refinement. In the subsequent sections, we outline on potential avenues for future studies, grounded in the primary insights gained from this dissertation.

Market Risk Chapter 2 indicates that incorporating information from an auxiliary significance level into a semiparametric risk model for joint (VaR, ES) can improve the forecasting accuracy of the risk model at the extreme level. However, our study only applies the new proposed models on data on energy futures due to their high volatility. One potential avenue for future research could be employing the A-GAS models for other asset classes with high volatility, such as crypto currencies, agricultural commodities, foreign exchange, and equities. Moreover, one could extend the proposed framework by estimating risk for more than two levels, or incorporating information from multiple auxiliary levels to optimize the performance at the extreme level. Also, our methodology can be utilized on alternative risk models such as various GARCH-type models, which originated from Bollerslev (1986).

Climate Risk Chapter 3 investigates the relationship between climate risk factors and the VaR and ES of equities, and proposes novel risk measures that capture the VaR and ES attributed to climate risk exposures. However, we only examine the impacts from climate transition risk factors in this thesis. Therefore, it would be natural to examine the effects of physical risk factors on VaR and ES, e.g., rising sea levels, wildfires, extreme temperature events, drought, flood, or hurricane-prone regions. Another compelling research direction could be to examine the impact of climate-related factors on other risk measures, such as volatility and expectile, and determine whether a similar relationship exists. Since our research on climate risk factors is conducted at the sector level, it is also worthwhile to investigate whether this relationship holds across different regions or countries, taking into account the local regulatory frameworks and environmental policies.

Credit Risk Chapter 4 delves into the relationship between corporate environmental performance and credit ratings in the US and EU. However, our study mostly focuses on well-developed countries, which have relatively mature environmental regulations and policies. Considering the rapid economic growth and environmental challenges in emerging economies, it might be insightful to study whether the positive relationship between credit ratings and environmental performance remains statistically significant in these regions. Clarkson et al. (2008) verify the positive relationship between corporate environmental performance and environmental disclosures. It would also be worthwhile to evaluate how the disclosure and transparency of firms about their environmental practices could influence their credit ratings. As businesses increasingly prioritize transparency and ESG performance, it would be valuable to explore how the social and governance factors of the firms influence their credit ratings.

References

Acerbi, C., Tasche, D., 2002. Expected Shortfall: A natural coherent alternative to Value at Risk. Economic Notes 31, 379–388.

Allen, S. L., 2012. Financial risk management: A practitioner's guide to managing market and credit risk. John Wiley & Sons.

Alsakka, R., Ap Gwilym, O., 2013. Rating agencies' signals during the european sovereign debt crisis: Market impact and spillovers. Journal of Economic Behavior & Organization 85, 144–162.

Alsakka, R., ap Gwilym, O., Vu, T. N., 2014. The sovereign-bank rating channel and rating agencies' downgrades during the european debt crisis. Journal of International Money and Finance 49, 235–257.

Andersson, M., Bolton, P., Samama, F., 2016. Hedging climate risk. Financial Analysts Journal 72, 13–32.

Anttila-Hughes, J., 2016. Financial market response to extreme events indicating climatic change. The European Physical Journal Special Topics 225, 527–538.

Artzner, P., 1997. Thinking coherently. Risk 10, 68–71.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9, 203–228.

Ashbaugh-Skaife, H., Collins, D. W., LaFond, R., 2006. The effects of corporate governance on firms' credit ratings. Journal of Accounting and Economics 42, 203–243.

Attig, N., El Ghoul, S., Guedhami, O., Suh, J., 2013. Corporate social responsibility and credit ratings. Journal of Business Ethics 117, 679–694.

Baghai, R. P., Servaes, H., Tamayo, A., 2014. Have rating agencies become more conservative? Implications for capital structure and debt pricing. The Journal of Finance 69, 1961–2005.

Baldauf, M., Garlappi, L., Yannelis, C., 2020. Does climate change affect real estate prices? Only if you believe in it. The Review of Financial Studies 33, 1256–1295.

Bansal, R., Kiku, D., Ochoa, M., 2016. Price of long-run temperature shifts in capital markets, working paper available at: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2827447`, accessed on Aug 2023.

Barnett, M., 2017. Climate change and uncertainty: An asset pricing perspective, working paper available at: `https://academic.oup.com/rfs/article/33/3/1024/5735312`, accessed on Aug 2023.

Barnett, M., Brock, W., Hansen, L. P., 2020. Pricing uncertainty induced by climate change. The Review of Financial Studies 33, 1024–1066.

Barney, J., 1991. Firm resources and sustained competitive advantage. Journal of Management 17, 99–120.

Bartram, S. M., Conrad, J., Lee, J., Subrahmanyam, M. G., 2022. Credit default swaps around the world. The Review of Financial Studies 35, 2464–2524.

Basel Committee on Banking Supervision, 2013. Fundamental review of the trading book: A revised market risk framework. Available at `https://www.bis.org/publ/bcbs265.pdf`, accessed on Aug 2023.

Bauer, R., Hann, D., 2010. Corporate environmental management and credit risk, working paper available at: `https://ssrn.com/abstract=1660470`, Accessed on Aug 2023.

Bernstein, A., Gustafson, M. T., Lewis, R., 2019. Disaster on the horizon: The price effect of sea level rise. Journal of Financial Economics 134, 253–272.

Best, R., Burke, P. J., Jotzo, F., 2020. Carbon pricing efficacy: Cross-country evidence. Environmental and Resource Economics 77, 69–94.

Bhandari, A., Golden, J., 2021. CEO political preference and credit ratings. Journal of Corporate Finance 68:101909.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bolton, P., Kacperczyk, M., 2021. Do investors care about carbon risk? Journal of Financial Economics 142, 517–549.

Bonsall IV, S. B., Holzman, E. R., Miller, B. P., 2017. Managerial ability and credit risk assessment. Management Science 63, 1425–1449.

Borghesi, R., Houston, J. F., Naranjo, A., 2014. Corporate socially responsible investments: CEO altruism, reputation, and shareholder interests. Journal of Corporate Finance 26, 164–181.

Burke, M., Hsiang, S. M., Miguel, E., 2015. Global non-linear effect of temperature on economic production. Nature 527, 235–239.

Cai, Y., Pan, C. H., Statman, M., 2016. Why do countries matter so much in corporate social performance? Journal of Corporate Finance 41, 591–609.

Chapman, K., Miller, G. S., White, H. D., 2019. Investor relations and information assimilation. The Accounting Review 94, 105–131.

Chava, S., 2014. Environmental externalities and cost of capital. Management Science 60, 2223–2247.

Chavez-Demoulin, V., Embrechts, P., Sardy, S., 2014. Extreme-quantile tracking for financial time series. Journal of Econometrics 181, 44–52.

Chen, J., 2018. On exactitude in financial regulation: Value-at-risk, expected shortfall, and expectiles. Risks 6, 61.

Chen, Q., Gerlach, R., Lu, Z., 2012. Bayesian Value-at-Risk and Expected Shortfall forecasting via the asymmetric Laplace distribution. Computational Statistics & Data Analysis 56, 3498–3516.

Choi, D., Gao, Z., Jiang, W., 2020. Attention to global warming. The Review of Financial Studies 33, 1112–1145.

Christensen, H. B., Hail, L., Leuz, C., 2021. Mandatory CSR and sustainability reporting: Economic analysis and literature review. Review of Accounting Studies 26, 1176–1248.

Christoffersen, P. F., 1998. Evaluating interval forecasts. International Economic Review pp. 841–862.

Clarkson, P. M., Li, Y., Richardson, G. D., Vasvari, F. P., 2008. Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. Accounting, Organizations and Society 33, 303–327.

Clarkson, P. M., Li, Y., Richardson, G. D., Vasvari, F. P., 2011. Does it really pay to be green? Determinants and consequences of proactive environmental strategies. Journal of Accounting and Public Policy 30, 122–144.

Cornaggia, K. J., Krishnan, G. V., Wang, C., 2017. Managerial ability and credit ratings. Contemporary Accounting Research 34, 2094–2122.

Cornish, E. A., Fisher, R. A., 1938. Moments and cumulants in the specification of distributions. Revue de l'Institut International de Statistique 5, 307–320.

Creal, D., Koopman, S. J., Lucas, A., 2013. Generalized autoregressive score models with applications. Journal of Applied Econometrics 28, 777–795.

Curtin, J., McInerney, C., Gallachóir, B. Ó., Hickey, C., Deane, P., Deeney, P., 2019. Quantifying stranding risk for fossil fuel assets and implications for renewable energy investment: A review of the literature. Renewable and Sustainable Energy Reviews 116:109402.

Danielsson, J., De Vries, C. G., 1998. Beyond the Sample: Extreme Quantile and Probability Estimation. Rotterdam Tinbergen Institute.

DeFond, M. L., Erkens, D. H., Zhang, J., 2016. Does PSM really eliminate the big n audit quality effect?, working paper available at: `https://ssrn.com/abstract=2472092`, Accessed on Aug 2023.

Del Brio, E. B., Mora-Valencia, A., Perote, J., 2020. Risk quantification for commodity ETFs: Backtesting value-at-risk and expected shortfall. International Review of Financial Analysis 70, 101163.

Dempster, M. A. H., 2002. Risk management: Value at risk and beyond. Cambridge University Press.

Diebold, F. X., Mariano, R. S., 2002. Comparing predictive accuracy. Journal of Business & Economic Statistics 20, 134–144.

Dietz, S., Bowen, A., Dixon, C., Gradwell, P., 2016. 'Climate Value at Risk' of global financial assets. Nature Climate Change 6, 676–679.

Dowd, K., 1998. Beyond Value at Risk: The new science of risk management, vol. 96. Wiley.

Duffie, D., Pan, J., 1997. An overview of Value at Risk. Journal of Derivatives 4, 7–49.

Dyck, A., Lins, K. V., Roth, L., Wagner, H. F., 2019. Do institutional investors drive corporate social responsibility? International evidence. Journal of Financial Economics 131, 693–714.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., Stroebel, J., 2020. Hedging climate change news. The Review of Financial Studies 33, 1184–1216.

Engle, R. F., Manganelli, S., 2004. CAViaR: Conditional autoregressive Value at Risk by regression quantiles. Journal of Business & Economic Statistics 22, 367–381.

European Commission, 2020. Summary report on the public consultation on the review of the non-financial reporting directive. Ref. Ares(2020)3997889 – 29/07/2020., European Commission. Available at `https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12129-revision-of-non-financial-reporting-directive/public-consultation_en`, Accessed on Aug 2023.

European Union, 2014. Directive 2014/95/EU, Official Journal of the European Union. Available at `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32014L0095`, Accessed on Aug 2023.

Fang, V. W., Tian, X., Tice, S., 2014. Does stock liquidity enhance or impede firm innovation? The Journal of Finance 69, 2085–2125.

Feldman, S. J., Soyka, P. A., Ameer, P. G., 1997. Does improving a firm's environmental management system and environmental performance result in a higher stock price? The Journal of Investing 6, 87–97.

Fernando, C. S., Sharfman, M. P., Uysal, V. B., 2017. Corporate environmental policy and shareholder value: Following the smart money. Journal of Financial and Quantitative Analysis 52, 2023–2051.

Ferreira, M. A., Gama, P. M., 2007. Does sovereign debt ratings news spill over to international stock markets? Journal of Banking & Finance 31, 3162–3182.

Fissler, T., Ziegel, J., 2016. Higher order elicitability and Osband's principle. Annals of Statistics 44, 1680–1707.

Fombrun, C., Shanley, M., 1990. What's in a name? Reputation building and corporate strategy. Academy of Management Journal 33, 233–258.

Freeman, R. E., 1984. Strategic management: A stakeholder approach. Cambridge University Press.

Ge, W., Liu, M., 2015. Corporate social responsibility and the cost of corporate bonds. Journal of Accounting and Public Policy 34, 597–624.

Gerlach, R., Wang, C., 2020. Semi-parametric dynamic asymmetric Laplace models for tail risk forecasting, incorporating realized measures. International Journal of Forecasting 36, 489–506.

Giglio, S., Kelly, B., Stroebel, J., 2021. Climate finance. Annual Review of Financial Economics 13, 15–36.

Gillan, S. L., Koch, A., Starks, L. T., 2021. Firms and social responsibility: A review of ESG and CSR research in corporate finance. Journal of Corporate Finance 66:101889.

Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.

Gneiting, T., 2011. Making and evaluating point forecasts. Journal of the American Statistical Association 106, 746–762.

Goldsmith-Pinkham, P., Gustafson, M., Lewis, R., Schwert, M., 2019. Sea level rise and municipal bond yields, Jacobs Levy Equity Management Center for Quantitative Financial Research Paper, Available at `https://dx.doi.org/10.2139/ssrn.3478364`, accessed on Aug 2023.

Gompers, P., Ishii, J., Metrick, A., 2003. Corporate governance and equity prices. The Quarterly Journal of Economics 118, 107–156.

Greening, D. W., Turban, D. B., 2000. Corporate social performance as a competitive advantage in attracting a quality workforce. Business & Society 39, 254–280.

Hainmueller, J., 2012. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. Political Analysis 20, 25–46.

Hainmueller, J., Xu, Y., 2013. Ebalance: A Stata package for entropy balancing. Journal of Statistical Software 54, 1–18.

Hansen, B. E., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.

Hart, J. D., 1994. Automated kernel smoothing of dependent data by using time series cross-validation. Journal of the Royal Statistical Society: Series B (Methodological) 56, 529–542.

Harvey, A. C., 2013. Dynamic Models for Volatility and Heavy Tails: With applications to Financial and Economic Time Series, vol. 52. Cambridge University Press.

Heckman, J. J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.

Hoepner, A. G., Oikonomou, I., Sautner, Z., Starks, L. T., Zhou, X., 2018. ESG shareholder engagement and downside risk, working paper available at: `https://ssrn.com/abstract=2874252`, Accessed on Aug 2023.

Hoga, Y., 2017. Testing for changes in (extreme) VaR. Econometrics Journal 20, 23–51.

Hoga, Y., 2021. The uncertainty in extreme risk forecasts from covariate-augmented volatility models. International Journal of Forecasting 37, 675–686.

Hong, H., Kubik, J. D., Scheinkman, J. A., 2012. Financial constraints on corporate goodness, working paper available at: `https://ssrn.com/abstract=1734164`, Accessed on Aug 2023.

Hong, H., Li, F. W., Xu, J., 2019. Climate risks and market efficiency. Journal of Econometrics 208, 265–281.

Hossain, A., Hossain, T., Jha, A., Mougoué, M., 2023. Credit ratings and social capital. Journal of Corporate Finance 78:102338.

Hsu, P.-h., Li, K., Tsou, C.-y., 2023. The pollution premium. The Journal of Finance 78, 1343–1392.

Hung, J.-C., Lee, M.-C., Liu, H.-C., 2008. Estimation of value-at-risk for energy commodities via fat-tailed GARCH models. Energy Economics 30, 1173–1191.

Huynh, T. D., Xia, Y., 2021. Climate change news risk and corporate bond returns. Journal of Financial and Quantitative Analysis 56, 1985–2009.

Hyndman, R. J., Athanasopoulos, G., 2018. Forecasting: Principles and Practice. OTexts.

Ilhan, E., Sautner, Z., Vilkov, G., 2021. Carbon tail risk. The Review of Financial Studies 34, 1540–1571.

Jagannathan, R., Ravikumar, A., Sammon, M., 2018. Environmental, social, and governance criteria: Why investors should care. Journal of Investment Management 16, 18–31.

Johnston, R., Jones, K., Manley, D., 2018. Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. Quality & Quantity 52, 1957–1976.

Jorion, P., 2000. Risk management lessons from long-term capital management. European Financial Management 6, 277–300.

Kai, L., Prabhala, N. R., 2007. Self-selection models in corporate finance. Handbook of Empirical Corporate Finance 1, 37–86.

Klassen, R. D., McLaughlin, C. P., 1996. The impact of environmental management on firm performance. Management Science 42, 1199–1214.

Koenker, R., 2004. Quantile regression for longitudinal data. Journal of Multivariate Analysis 91, 74–89.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50.

Konar, S., Cohen, M. A., 2001. Does the market value environmental performance? Review of Economics and Statistics 83, 281–289.

Kourouma, L., Dupre, D., Sanfilippo, G., Taramasco, O., 2010. Extreme Value at Risk and Expected Shortfall during financial crisis, working paper available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1744091, accessed on Aug 2023.

Kumar, A., Xin, W., Zhang, C., 2019. Climate sensitivity, mispricing, and predictable returns, working paper available at https://dx.doi.org/10.2139/ssrn.3331872, accessed on Aug 2023.

Kupiec, P. H., 1995. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives 3, 73–84.

Lahouel, B. B., Zaied, Y. B., Managi, S., Taleb, L., 2022. Re-thinking about u: The relevance of regime-switching model in the relationship between environmental corporate social responsibility and financial performance. Journal of Business Research 140, 498–519.

Lawrence, A., Minutti-Meza, M., Zhang, P., 2011. Can big 4 versus non-big 4 differences in audit-quality proxies be attributed to client characteristics? The Accounting Review 86, 259–286.

Lesk, C., Rowhani, P., Ramankutty, N., 2016. Influence of extreme weather disasters on global crop production. Nature 529, 84–87.

Li, H., Liu, Z., Wang, S., 2022. Vines climbing higher: Risk management for commodity futures markets using a regular vine copula approach. International Journal of Finance & Economics 27, 2438–2457.

Liang, H., Renneboog, L., 2017. On the foundations of corporate social responsibility. The Journal of Finance 72, 853–910.

Liu, H.-C., Hung, J.-C., 2010. Forecasting S&P100 stock index volatility: The role of volatility asymmetry and distributional assumption in GARCH models. Expert Systems with Applications 37, 4928–4934.

Maciel, L., 2021. Cryptocurrencies Value-at-Risk and Expected Shortfall: Do regime-switching volatility models improve forecasting? International Journal of Finance & Economics 26, 4840–4855.

McNeil, A. J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of empirical finance 7, 271–300.

Menz, K.-M., 2010. Corporate social responsibility: Is it rewarded by the corporate bond market? A critical note. Journal of Business Ethics 96, 117–134.

Merton, R. C., 1974. On the pricing of corporate debt: The risk structure of interest rates. The Journal of Finance 29, 449–470.

Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347–370.

Ng, A. C., Rezaee, Z., 2015. Business sustainability performance and cost of equity capital. Journal of Corporate Finance 34, 128–149.

Nolde, N., Ziegel, J. F., 2017. Elicitability and backtesting: Perspectives for banking regulation. The Annals of Applied Statistics 11, 1833–1874.

Oikonomou, I., Brooks, C., Pavelin, S., 2014. The effects of corporate social performance on the cost of corporate debt and credit ratings. Financial Review 49, 49–75.

Oster, E., 2019. Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics 37, 187–204.

Painter, M., 2020. An inconvenient cost: The effects of climate change on municipal bonds. Journal of Financial Economics 135, 468–482.

Paris Agreement, 2015. Climate policies in the EU and USA. Different approaches, convergent outcomes?, European Parliamentary Research Service. Available at `https://www.europarl.europa.eu/RegData/etudes/`

BRIE/2015/571347/EPRS_BRI%282015%29571347_EN.pdf, Accessed on Aug 2023.

Patton, A. J., Ziegel, J. F., Chen, R., 2019. Dynamic semiparametric models for Expected Shortfall (and value-at-risk). Journal of Econometrics 211, 388–413.

Roccioletti, S., 2015. Backtesting Value at Risk and Expected Shortfall. Springer.

Rosenbaum, P. R., Rubin, D. B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.

Russo, M. V., Fouts, P. A., 1997. A resource-based perspective on corporate environmental performance and profitability. Academy of Management Journal 40, 534–559.

Sadorsky, P., 1999. Oil price shocks and stock market activity. Energy economics 21, 449–469.

Safiullah, M., Kabir, M. N., Miah, M. D., 2021. Carbon emissions and credit ratings. Energy Economics 100:105330.

Seltzer, L. H., Starks, L., Zhu, Q., 2022. Climate regulatory risk and corporate bonds, working paper available at: https://ssrn.com/abstract=3563271, Accessed on Aug 2023.

Sharfman, M. P., Fernando, C. S., 2008. Environmental risk management and the cost of capital. Strategic Management Journal 29, 569–592.

Shipman, J. E., Swanquist, Q. T., Whited, R. L., 2017. Propensity score matching in accounting research. The Accounting Review 92, 213–244.

Stellner, C., Klein, C., Zwergel, B., 2015. Corporate social responsibility and eurozone corporate bonds: The moderating role of country sustainability. Journal of Banking & Finance 59, 538–549.

Storti, G., Wang, C., 2022. Nonparametric Expected Shortfall forecasting incorporating weighted quantiles. International Journal of Forecasting 38, 224–239.

Tang, D. Y., Zhang, Y., 2020. Do shareholders benefit from green bonds? Journal of Corporate Finance 61:101427.

Tang, T. T., 2009. Information asymmetry and firms' credit market access: Evidence from Moody's credit rating format refinement. Journal of Financial Economics 93, 325–351.

Tankov, P., Tantet, A., 2019. Climate data for physical risk assessment in finance, working paper available at: `https://dx.doi.org/10.2139/ssrn.3480156`, Accessed on Aug 2023.

Taylor, J. W., 2019. Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. Journal of Business & Economic Statistics 37, 121–133.

Taylor, J. W., 2020. Forecast combinations for Value at Risk and Expected Shortfall. International Journal of Forecasting 36, 428–441.

Trinh, V. Q., Cao, N. D., Li, T., Elnahass, M., 2023. Social capital, trust, and bank tail risk: The value of ESG rating and the effects of crisis shocks. Journal of International Financial Markets, Institutions and Money 83, 101740.

Turban, D. B., Greening, D. W., 1997. Corporate social performance and organizational attractiveness to prospective employees. Academy of Management Journal 40, 658–672.

Vassalou, M., Xing, Y., 2004. Default risk in equity returns. The journal of Finance 59, 831–868.

Waddock, S. A., Graves, S. B., 1997. The corporate social performance–financial performance link. Strategic Management Journal 18, 303–319.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.

White, H., Kim, T.-H., Manganelli, S., 2015. VAR for VaR: Measuring tail dependence using multivariate regression quantiles. Journal of Econometrics 187, 169–188.

Wilde, J. H., 2017. The deterrent effect of employee whistleblowing on firms' financial misreporting and tax aggressiveness. The Accounting Review 92, 247–280.

Ziegel, J. F., 2016. Coherence and elicitability. Mathematical Finance 26, 901–918.