Essays on Corporate Social Responsibility — Performance, Risk, and Value

D I S S E R T A T I O N

of the University of St.Gallen,

School of Management,

Economics, Law, Social Sciences,

International Affairs and Computer Science,

to obtain the title of

Doctor of Philosophy in Management

submitted by

Daniel Valentin Fauser

from

Germany

Approved on the application of

Prof. Dr. Andreas Grüner

and

Prof. Dr. Tami Dinh

Dissertation no. 5043

Difo-Druck GmbH, Untersiemau 2021

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St. Gallen, October 23, 2020

The President:

Prof. Dr. Bernhard Ehrenzeller

To my lovely girlfriend, Josepha, who always stood by my side and sincerely supported me with her encouragement, critical counsel, patience, and love in challenging times.

To my parents who were always there for me and who made this endeavor possible for me in the first place. Thank you for your unconditional love, support, and care.

To Sebastian, my esteemed co-author and friend, who was always interested in my ideas and open to in-depth discussions. Thank you so much for working hard together on our project and for enriching my thesis on such an unprecedented scale.

To Ulrich, my experienced mentor, who was always open to my concerns and who helped me with his valuable experience and advice to make the right decisions.

To Robert, one of my early co-authors and friend, who saw and believed in my potential and thus supported me intellectually and professionally.

To my friends from the two boards of DocNet (18/19, 19/20), Anita, Anna-Sophia, Michael, Felix, Kathrin, Florian, Ben, Charlotte, Julian, Raphael, Christian, Sabrina, Babsi, Marcel, Lea, Natalia, and Eddy who made life in St.Gallen such an incredible experience.

Carpe diem!

Acknowledgements

I am grateful to Prof. Dr. Tami Dinh and Prof. Dr. Andreas Grüner for their continued support and the very valuable feedback I have received from them during the whole process of my dissertation.

Paper 1: I am very thankful for the helpful comments from Prof. Jaap Bos, Ph.D., and Prof. Paul Smeets, Ph.D., at the conference "Managing and Financing Responsible Businesses," the inaugural conference of the Global Research Alliance for Sustainable Finance and Investment (GRASFI), in Maastricht from September 5th to 7th, 2018, and from the conference participants at the ACA Symposium at the University of St.Gallen, from November 7th to 9th, 2018. I am especially thankful to Prof. Dr. Frank Verbeeten from the University of Amsterdam for his extensive feedback during the ACA Symposium. I further thank two anonymous reviewers of the Journal of Environmental Law & Policy for their comments.

Paper 2: My co-author and I are grateful to Prof. Dr. Bert Scholtens, Prof. Dr. Timo Busch, Prof. Dr. Judith Walls, and the seminar participants at Maastricht University, University of Liechtenstein, the Annual GRASFI Conference in Oxford from September 3rd to 5th, 2019, and the Corporate Finance Day in Groningen from September 18th to 20th, 2019, for their helpful feedback and comments. We further thank the executive editor of the Financial Analysts Journal, Prof. Stephen J. Brown, Ph.D., and two anonymous reviewers for their comments on how to improve and tailor our paper to the journal's audience.

Paper 3: My co-author and I very much appreciate the continuous feedback and support of Prof. Dr. Sebastian Utz and Prof. Dr. Jeroen Derwall during the development of this paper. We further thank an anonymous reviewer for his valuable comments.

Zürich, December 2020 Daniel V. Fauser

Abstract

This thesis is dedicated to the relationship between corporate social performance (CSP) and corporate financial performance (CFP). More specifically, it identifies and analyzes three distinct accounting- and finance-related topics from an investor's perspective: The first paper tests whether the materiality of information on CSP plays a role in explaining earnings quality. Preliminary results show that firms performing well on material CSP have higher accruals quality. In contrast, material and non-material CSP is linked to higher real earnings management. Robustness checks show that industry regulation plays a major moderating role, whereas materiality of CSP is not necessarily a reliable indicator of higher earnings quality.

The second paper investigates whether CSP is linked to the risk of facing litigation and whether CSP provides an insurance-like effect against litigation. The results show that an improvement in the measure of negative CSP of one standard deviation of an average sample firm reduces litigation risk from 3.1% to 2.4%. The losses in market value from litigation for a firm with low CSP are twice as high as the losses for a firm with high CSP. Implementing a trading strategy that is based on these findings yielded monthly positive alphas of 1.31%.

The third paper examines whether CSP increases the prediction accuracy for credit risk in various machine learning algorithms. It finds that the prediction accuracy varies considerably between algorithms, with advanced machine learning algorithms (e.g., random forests) outperforming traditional ones (e.g., linear regression). This finding is probably best explained by the ability of advanced machine learning algorithms to capture non-linearity and complex interaction effects in the data. Moreover, including information on CSP into firm credit risk prediction does not consistently increase prediction accuracy.

All three papers converge on the relevance of CSP for the financial performance of firms. The first paper focuses on an accounting perspective on the relationship between CSP and CFP. The second paper introduces the market perspective. The third paper employs a rigorous predictive approach. The three papers have in common that they provide relevant and rigorous guidance on fundamental drivers of the CSP–CFP relationship to investors and analysts.

Abstract - German

Diese Dissertation ist der Beziehung zwischen der “Corporate Social Performance” (CSP) und der finanziellen Performance von Unternehmen (“Corporate Financial Performance”, CFP) gewidmet. Im Einzelnen werden drei verschiedene rechnungslegungs- und finanzbezogene Themen identifiziert und aus der Investorenperspektive untersucht: Im ersten Paper wird getestet, ob die Wesentlichkeit von Informationen über CSP eine Rolle bei der Erklärung der Ertragsqualität spielt. Erste Ergebnisse zeigen, dass Firmen, die bei der wesentlichen CSP gut abschneiden, eine höhere Rückstellungsqualität (engl., “accruals quality”) aufweisen. Im Gegensatz dazu sind wesentliche und nicht wesentliche CSP mit einer höheren Manipulation von realen Erträgen (engl., “real earnings management”) verbunden. Tests zur Robustheit der Ergebnisse zeigen, dass die Regulierung der Industrie eine wichtige moderierende Rolle spielt, während die Wesentlichkeit von CSP nicht zwingend ein zuverlässiger Indikator für eine höhere Ertragsqualität ist.

Das zweite Paper untersucht, ob CSP mit dem Risiko eines Rechtsstreits verbunden ist und ob CSP einen versicherungsähnlichen Effekt gegen Rechtsstreitigkeiten bietet. Es zeigt sich, dass eine Verbesserung des Maßes der negativen CSP um eine Standardabweichung einer durchschnittlichen Stichprobenfirma das Risiko eines Rechtsstreits von 3,1% auf 2,4% reduziert. Die Marktwertverluste aus Rechtsstreitigkeiten, die Firmen mit niedriger CSP aufweisen, sind ausserdem doppelt so hoch wie die Verluste von Firmen mit hohem CSP. Die Umsetzung einer Handelsstrategie, die auf diesen Ergebnissen basiert, lieferte monatliche positive Alphas von 1,31%.

Das dritte Paper untersucht, ob CSP die Vorhersagegenauigkeit für das Kreditrisiko unter Verwendung verschiedener Algorithmen des maschinellen Lernens erhöht. Die Vorhersagegenauigkeit zwischen den verschiedenen Algorithmen variiert erheblich, wobei fortgeschrittene Algorithmen des maschinellen Lernens (z.B., “random forests”) die traditionellen Algorithmen (z.B., lineare Regression) übertreffen. Dieses Ergebnis lässt sich wahrscheinlich am besten durch die Fähigkeit dieser fortgeschrittenen Algorithmen erklären, Nichtlinearität und komplexe Interaktionseffekte in den Daten zu erfassen. Darüber hinaus erhöht die Einbeziehung von Informationen über CSP in die Kreditrisikovorhersage nicht durchgängig die Vorhersage-

genauigkeit.

Alle drei Beiträge handeln von der Relevanz der CSP für die finanzielle Leistungsfähigkeit von Unternehmen. Der erste Beitrag konzentriert sich auf die Rechnungslegungsperspektive hinsichtlich der Beziehung zwischen CSP und CFP. Der zweite Beitrag führt in die Marktperspektive ein. Im dritten Beitrag wird ein rigoroser, prognostischer Ansatz angewendet. Den drei Beiträgen ist unter anderem gemeinsam, dass sie Investoren und Analysten relevante und präzise Orientierungshilfen zu grundlegenden Faktoren der Beziehung zwischen CSP und CFP bieten.

Table of Contents

List of Figures **xii**

List of Tables **xiv**

List of Abbreviations **xviii**

I Thematic Context and Motivation **1**

I.1 Derivation of Corporate Social Responsibility and Corporate Social Performance 4

I.2 Main Terminology 6

I.3 Review of Selected Empirical Literature 7

I.4 Research Ideas and Objectives 13

I.4.1 Corporate Social Performance, Earnings Quality, and Materiality . . . 13

I.4.2 Corporate Social Performance and Litigation Risk 15

I.4.3 Corporate Social Performance and Credit Risk Prediction 16

I.4.4 Similarities and Differences Between the Papers 17

I.5 Overview of the Data and the Research Design 23

I.5.1 1st Paper: European Data and Multiple Linear Panel Regression 23

I.5.2 2nd Paper: US Data, Multiple Probit Panel Regression, Event Study Methodology, and Portfolio Approach 24

I.5.3 3rd Paper: Global Data and Machine Learning Regression Design . . . 25

I.6 Structure 29

I.7 Data and Code Repository 29

I.8 Publication Status and Strategy 30

II Paper 1: Corporate Social Performance and Earnings Quality: The Role of Materiality **31**

II.1 Introduction 32

II.2 Theoretical Background, Literature Review, and Hypothesis Development . . . 34

II.2.1 Corporate Social Performance and Materiality 34

II.2.2 Empirical Literature on Earnings Quality 35

II.2.3 Hypotheses Development 39

II.3 Methodology & Data . . . 41
II.3.1 ESG Data . . . 42
II.3.2 Materiality Classification Approach of ESG Issues . . . 44
II.3.3 Recalculation of the Materiality-Based ESG Scores . . . 45
II.3.4 Proxies for Earnings Quality . . . 46
II.3.5 Empirical Model and Control Variables . . . 48
II.4 Empirical Results . . . 50
II.4.1 Descriptive Statistics . . . 50
II.4.2 The Relation between Material and Non-Material CSP and Accruals Quality . . . 54
II.4.3 The Relation between Material and Non-Material CSP and Real Earnings Management . . . 56
II.4.4 Additional Analyses and Robustness Checks . . . 61
II.5 Concluding Remarks . . . 64

III Paper 2: Risk Mitigation of Corporate Social Performance in US Class Action Lawsuits 67

III.1 Introduction . . . 68
III.2 Corporate Social Performance and the Probability of Litigation . . . 72
III.3 The Impact of Litigation on Firm Value . . . 73
III.4 The Insurance-Like Effect of Corporate Social Performance During Litigation . 74
III.5 Testing the Relation between CSP and Litigation Risk . . . 75
III.5.1 Measuring Corporate Social Performance . . . 75
III.5.2 Proxies for Litigation Risk . . . 76
III.5.3 Estimating the Impact of CSP on Litigation Risk . . . 76
III.5.4 ESG Data . . . 78
III.5.5 Class Action Lawsuit Data . . . 79
III.5.6 Descriptive Statistics . . . 80
III.5.7 Estimating Litigation Risk Using Probit Regressions . . . 82
III.6 Testing the Insurance-Like Effect in Class Action Lawsuits . . . 85
III.6.1 Market-Based Estimation of the Insurance-Like Effect . . . 86
III.6.2 Subsample on Class Action Lawsuit Filings . . . 87

III.6.3 Evidence on the Insurance-Like Effect of Corporate Social Performance 87
III.6.4 Robustness Checks . . . 89
III.6.5 Cross-Sectional Analysis After Applying Propensity Score Matching . 90
III.7 Implementation of the Findings in a Trading Strategy . . . 94
III.7.1 Factor Portfolios and Abnormal Risk-Adjusted Returns . . . 95
III.7.2 Transaction Costs of Factor Portfolios . . . 99
III.8 Concluding Remarks and Practical Implications . . . 100

IV Paper 3: Corporate Social Irresponsibility and Credit Risk Prediction: A Machine Learning Approach 102

IV.1 Introduction . . . 103
IV.2 Literature Review and Framework . . . 107
IV.2.1 Machine Learning Applications . . . 107
IV.2.2 Corporate Social Irresponsibility and Credit Risk . . . 108
IV.3 Research Design . . . 111
IV.3.1 Predictors for Corporate Social Irresponsibility . . . 111
IV.3.2 Additional Accounting- and Market-Related Predictors . . . 113
IV.3.3 Proxy for Credit Risk . . . 114
IV.3.4 Sample Development . . . 115
IV.3.5 Prediction Using a Selection of Machine Learning Algorithms . . . 116
IV.4 Descriptive Statistics . . . 123
IV.5 Empirical Results . . . 125
IV.5.1 Main Results . . . 125
IV.5.2 Additional Analyses – Regression Design . . . 133
IV.5.3 Predicting Investment Grade vs. Non-Investment Grade – Classification Design . . . 136
IV.6 Contributions to Research and Practical Implications . . . 140
IV.7 Concluding Remarks . . . 142

V Overall Concluding Remarks 144

V.1 Main Results and Contribution to the Seminal Literature . . . 144
V.2 Note on a More General Limitation of Studies on CSP . . . 149

V.3 Avenues for Future Research 150

References **153**

A Appendix: Paper 1 **172**

A.1 Materiality Guidance of the Sustainability Accounting Standards Board 172

A.2 Yearly Sample Distribution 173

A.3 Overview of the TR ESG Score Recalculation Procedure 174

A.4 Derivation of the Modified Accruals Quality Model 174

A.5 Derivation of the Real Earnings Management (REM) Models 176

A.6 Interpretation of the Accruals Quality Proxy 177

A.7 Interpretation of the Real Earnings Management Proxies 178

A.8 Robustness: Additional Analysis using Positive and Negative AQ 179

B Appendix: Paper 2 **180**

B.1 Variable Definitions 180

B.2 Examples for Class Action Lawsuits 181

B.3 Distribution of Sample Firms Across Sectors 184

B.4 Firm Percentages Facing Litigation 184

B.5 Distribution of Class Action Lawsuits per Year 185

B.6 Robustness: Logit – Estimation of Litigation Risk 186

B.7 Robustness: Baseline Estimation of Litigation Risk 187

B.8 Detailed Cumulative Abnormal Returns 188

B.9 Mean Abnormal Returns – Tertiles 189

B.10 Robustness: Alternative Models for Estimating Abnormal Returns 190

B.11 Robustness: Alternative Estimation Window 191

B.12 Robustness: Alternative Event Window 192

B.13 Robustness: Alternative Cut-Offs 193

B.14 Robustness: Placebo Event Study 194

B.15 Robustness: Winsorized CAR 195

B.16 Mean CAR per Industry 196

B.17 Robustness: Monthly Fama-French 5-Factor Alphas 196

C Appendix: Paper 3 **197**

C.1 Specification of Optimization Problems 197

C.2 Variable Definitions 203

C.3 Algorithm Default Hyper-Parameters 204

C.4 Robustness: US Data 205

C.5 Additional Analysis: ESG Issues with High Reach and High Severity 207

C.6 Robustness: Increased Test Sample Size 208

C.7 Robustness: Alternative Specification 210

C.8 Robustness: Non-Stratified Samples 213

C.9 Additional Analysis: Four-Quarter Lag 216

C.10 Additional Analysis: Post-2015 Data 218

C.11 Additional Analysis: Moody's and Fitch Credit Rating Data 220

List of Figures

Fig. 1: Sustainable Development Goals of the United Nations 3
Fig. 2: Main Terminology 7
Fig. 3: Mean Cumulative Abnormal Returns – Market Approach 89
Fig. 4: Mean Cumulative Abnormal Returns – Propensity Score Matching . . . 92
Fig. 5: Graphical Approach 122
Fig. 6: Statistical Significance of Prediction Accuracy – Baseline Model 129
Fig. 7: Model Prediction Accuracy – Peak RRI 134
Fig. A.1: Step-by-Step Calculation of TR ESG Scores 174
Fig. A.2: Interpreting the Sign of Accruals Quality 177
Fig. A.3: Interpreting the Sign of Real Earnings Management 178
Fig. A.4: Interpreting the Sign of Real Earnings Management (2) 178
Fig. B.1: Mean Abnormal Returns 189
Fig. B.2: Alternative Estimation Window 191
Fig. B.3: Alternative Event Window 192
Fig. B.4: Alternative Cut-Offs 193
Fig. B.5: Placebo Event Study 194
Fig. B.6: Winsorized CAR 195

List of Tables

Tab. 1: Selected Literature on Corporate Social Responsibility and Performance . 10
Tab. 2: Thesis Overview 20
Tab. 3: Overview of the Data, Methodology, and Test Statistics 28
Tab. 4: Summary Statistics 52
Tab. 5: Pairwise Correlations 53
Tab. 6: Regressing Accruals Quality on Material and Non-Material CSP 55
Tab. 7: Regressing Real Earnings Management on Overall CSP 58
Tab. 8: Regressing Real Earnings Management on Material and Non-Material CSP 59
Tab. 9: Analysis of Sub-Samples of Firms from Regulated and Non-Regulated Industries 63
Tab. 10: Summary Statistics 81
Tab. 11: Estimation of Litigation Risk in Probit Regressions 83
Tab. 12: Cross-Sectional Regression of $CAR_{t=10}$ 94
Tab. 13: Long-only Trading Strategies - Monthly Fama-French 5-Factor + Momentum Alphas 98
Tab. 14: Summary Statistics – Predictors and Target 124
Tab. 15: Summary Statistics – ESG Issues 126
Tab. 16: Prediction Accuracy of Machine Learning Algorithms – Baseline Model 128
Tab. 17: Prediction Accuracy – Comparison of Alternative Models 131
Tab. 18: Additional Analysis – Investment Grade vs. Non-Investment Grade . . . 139
Tab. 19: Main Findings and Contributions 148
Tab. A.1: Number of Observations per Industry and Year 173
Tab. A.2: Robustness Check – Regressing Positive and Negative AQ on Material and Non-Material CSP 179
Tab. B.1: Variable Definitions 180
Tab. B.2: Detailed Information on Examples for Class Action Lawsuits 181
Tab. B.3: Distribution of Sample Firms Across Sectors 184
Tab. B.4: Firm Percentages Facing Litigation 184
Tab. B.5: Distribution of Class Action Lawsuits per Year 185
Tab. B.6: Robustness Check – Estimation of Litigation Risk in Logit Regressions . 186

Tab. B.7: Baseline Estimation of Litigation Risk in Multiple Panel Regressions . . 187
Tab. B.8: Cumulative Abnormal Returns 188
Tab. B.9: Alternative Models for the Event Study Estimation of Abnormal Returns 190
Tab. B.10: Mean $CAR_{t=10}$ per Industry 196
Tab. B.11: Robustness: Base Trading Strategies – Monthly Fama-French 5-Factor Alphas 196
Tab. C.12: Variable Definitions 203
Tab. C.13: Algorithm Default Hyper-Parameters 204
Tab. C.14: Robustness Check – US Data 205
Tab. C.15: Additional Analysis – ESG Issues with High Reach and High Severity . . 207
Tab. C.16: Robustness Check – Increased Test Size 208
Tab. C.17: Robustness Check – Alternative Specification 211
Tab. C.18: Robustness Check – Non-Stratified Samples 214
Tab. C.19: Additional Analysis – Four Quarters Lag 216
Tab. C.20: Additional Analysis – Post-2015 Subsample 218
Tab. C.21: Additional Analysis – Moody's Credit Rating Data 221
Tab. C.22: Additional Analysis – Fitch Credit Rating Data 223

List of Abbreviations

ACFO Abnormal Operating Cash Flow

AdaBoostC AdaBoost Classification

AdaBoostR AdaBoost Regression

ADIEXP Abnormal Discretionary Expense

API Application Programming Interface

APROD Abnormal Production Cost

AQ Accruals Quality

AQM Accruals Quality Model

AR Abnormal Return

CAR Cumulative Abnormal Return

CDS Credit Default Swap

CFO Operating Cash Flow

CFP Corporate Financial Performance

CO_2 Carbon Dioxide

COGS Cost of Goods Sold

COMBREM Combined Measure of REM

CRSP The Center for Research in Security Prices

CSI Corporate Social Irresponsibility

CSP Corporate Social Performance

CSR Corporate Social Responsibility

ΔINV Change in Inventory

DTC Decision Tree Classification

DTR Decision Tree Regression

EQ Earnings Quality

ER Expected Return

ESG Environmental, Social, and Governance

FE Fixed Effects

FN False Negative

FP False Positive

FPS Francis, Philbrick, and Schipper

GBC Gradient Boosting Classification

GBR Gradient Boosting Regression

IDE Integrated Development Environment

IPO Initial Public Offering

KLD Kinder, Lydenburg, and Domini

KNNC K-nearest Neighbors Classification

KNNR K-nearest Neighbors Regression

ME Marginal Effects

MKTB Market-to-Book Ratio

ML Machine Learning

MLPC Multi-layer Perceptron Classification

MLPR Multi-layer Perceptron Regression

N Number of Observations

NGO Non-governmental Organization

OLS Ordinary Least Squares

PPE Property, Plant, and Equipment

PRI Principles for Responsible Investments

PSM Propensity Score Matching

R Actual Return

RBF Radial Basis Function

R&D Research & Development

REM Real Earnings Management

ΔREV Change in Revenue

RFC Random Forest Classification

RFR Random Forest Regression

RMSE Root Mean Squared Error

ROA Return on Assets

ROE Return on Equity

RRI RepRisk Index

SASB Sustainability Accounting Standards Board

SCAC Securities Class Action Clearinghouse

SD Standard Deviation

SDGs Sustainable Development Goals

SG&A Selling, General & Administrative Expense

S&P Standard & Poor's

SVC Support Vector Classification

SVR Support Vector Regression

TA Total Assets

TCA Total Current Accruals

TP True Positive

US United States

USD United States Dollar

WRDS Wharton Research Data Services

I Thematic Context and Motivation

Just before the end of the second decade of the 21st century, the Fridays for Future movement suddenly struck a blow in politics. The campaign was considered as the onset of a global and all-embracing development that is capable of changing the political and social landscape fundamentally. Finally, environmental sustainability was on the right track to becoming mainstream, even though scientists already long ago emphasized the ecological degradation and resulting social inequality caused by humans. It is striking that often there is such a strong emphasis on individual behavior. What is frequently overlooked—or at least neglected—is that the absolute majority of greenhouse gas emissions originates from the business activities of multinational corporations. In fact, as a recent report by the world-renowned Climate Accountability Institute showed, one-third of all carbon emissions since 1965 is attributable to an astonishing number of 20 firms, all from the fossil fuel industry (Taylor and Watts, 2019). Let alone the vast sums these companies spend on lobbying to keep everything as is or to influence the regulatory environment in their favor.

However, it is not only the soaring carbon emissions of multinational firms that led to the increasing importance of the Fridays for Future movement and the role corporations play therein. Environmental degradation and the resulting societal inequality is evident in many areas, some of which are already more visible than others. The examples for potential ecological disasters are numerous: Degrading coral reefs due to water pollution, rising sea levels due to the meltdown of glaciers, climate refugees due to the loss of livelihoods, and more frequent droughts and floods especially in poorer regions of the world. Most of these developments are attributable to business activities to a certain extent. According to the United States Environmental Protection Agency, the threats to coral reefs are plentiful. For instance, coral reefs undergo physical damage or destruction from fishing practices (overfishing) and gear, recreational misuse, dredging, and quarrying but also pollution that originates on lands such as trash and micro-plastics, pathogens, or toxic substances that make their way into coastal waters.[1] While impacts of business activities on coral reefs might be more subtle, the role of corporations in causing the sea levels to rise and the ascending average temperature due to pollution is

[1] For more information, please see https://www.epa.gov/coral-reefs/threats-coral-reefs

more prominent.

Besides the environmental impact of corporations, business activities always have a social component since they require human workforce. One might be tempted to say that a business without people is no business. At the forefront of this social dimension are probably the customers and employees of a business. However, various reports from non-governmental organizations and governmental agencies frequently raise concerns about issues such as the gender pay gap, lack of access to health care, hazardous products, and human rights violations. In the digital era, where pictures and news spread across the world in seconds, corporations can no longer ignore that they are intertwined with various stakeholders and that they increasingly face the (regulatory) necessity to address the issues around their social impact. Given these developments, it is reasonable to conclude that the importance of business activities in environmental, social, and governmental matters will very likely expand. Academic research plays an important role in examining the consequences of the way society is dealing with upcoming issues in that area.

In politics, developments like these have not only contributed to the emergence of movements like Fridays for Future but are also pushing politicians to sign global agreements to which all nations agree to adhere. The most famous agreement, the Paris Agreement on climate, set some targets that all involved parties (i.e., mainly states) agreed upon. For instance, the central aim is to take action to limit the global temperature increase to two degrees Celsius compared to pre-industrial levels.[2] Reaching this ambitious but possible objective requires politics to ensure that corporations and businesses are significantly contributing to this target, eventually.

The year 2015 was not only characterized by the Paris Agreement, but also by the 2030 Agenda on Sustainable Development and its 17 Sustainable Development Goals (SDGs) of the United Nations (see Figure 1). This agenda is supposed to be a "shared blueprint for peace and prosperity for people and the planet, now and into the future."[3] The 17 SDGs address many societal and environmental challenges, the mastering of which requires a considerable effort from states and—most importantly—corporations. The achievement of goals such as "good health

[2]For more information on the Paris Agreement, please see https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

[3]For more information on the SDGs, please see: https://sustainabledevelopment.un.org/sdgs

and well-being," "clean water and sanitation," "decent work and economic growth," "sustainable cities and communities," or "responsible consumption and production" would be unimaginable without the help of corporations. Many corporations have recognized the urgency and importance of the challenges and therefore committed to helping to achieve the SDGs. On February 17th, 2020, Jeff Bezos, Amazon's CEO and richest man in the world, personally pledged $10 billion to address the climate crisis, as The New York Times reported.[4] Philanthropic pledges like these underline the urgency and significance of the environmental and societal challenges, despite the legitimate criticism of this kind of individual philanthropy.

Figure 1: Sustainable Development Goals of the United Nations



In light of these developments, this thesis intends to answer some of the questions that arise with the growing importance of environmental, social, and governance (ESG) issues in the business world. In the next chapter, I will discuss the motivation and relevance of this thesis from the perspective of contemporary and seminal literature in more detail. In a sense, I will go back a step, back into the year 1970.

[4] Please see the following link for the report: https://www.nytimes.com/2020/02/17/technology/jeff-bezos-climate-change-earth-fund.html

I.1 Derivation of Corporate Social Responsibility and Corporate Social Performance

In academic literature, the debate about the environmental and societal role of businesses first got real momentum in the 1970s. Milton Friedman famously described the social responsibility of business is to increase its profits, sometimes also better known as the phrase "the business of business is business" (Friedman, 1970). According to Friedman (1970), only individuals can have responsibilities and not corporations. He also argued that those individuals steering the corporations, corporate executives, are solely responsible to their employers, i.e., the owners of the companies: the shareholders. He further states that, usually, shareholders' interests are to maximize profits within the rules of law and ethical custom (Friedman, 1970). The notion of shareholder value maximization dominated and partly still dominates the common understanding of business. However, this shareholder value notion was constantly challenged by alternative perceptions of corporate responsibility that were discussed under many different names (e.g., corporate social responsibility (CSR), stakeholder approach, good corporate citizenship, etc.).

Even though ideas of CSR[5] and sustainability evolved earlier than 1979, Carroll (1979) was the first to develop a systematic framework that defines CSR as going beyond the notion of maximizing shareholder value. Therefore, Carroll (1979) divided CSR into four categories, which have strongly influenced the idea of overall corporate responsibility: economic responsibilities, legal responsibilities, ethical responsibilities, and discretionary (or philanthropic) responsibilities. While Carroll (1979) did not conceive any category as superior to another, he suggested that, in the early history of business, the economic and legal categories received more emphasis compared to the ethical and discretionary ones. These corporate responsibilities were the foundation for the development of a systematic framework for corporate social performance (CSP).

Carroll (1979) ascribed the following three aspects to the term CSP: First, the identification and assessment of a firm's responsibilities (see the above definition of CSR); Second, the determination which CSR issues are relevant (e.g., product safety, human rights, emissions,

[5] In the early days of CSR, the term only addressed the social dimension of business. Sustainability considerations came in later (around the 1990s) and led to the convergence of CSR and environmental sustainability into the single umbrella term CSR.

etc.); And third, the specification of the philosophy of response (so-called "social responsiveness"). Social responsiveness can be either reactive, defensive, accommodative, or proactive (Carroll, 1979). This integrated framework of CSP was the first approach to bring together social responsibilities and social responsiveness (i.e., the realization of the social responsibilities) (Wartick and Cochran, 1985). The two dimensions were then complemented by a third one, social issues management, which is the method to operationalize social responsiveness (Wartick and Cochran, 1985). This operationalization of social responsiveness includes policies for issues identification, issues analysis, and response development (Wartick and Cochran, 1985).

Besides Carroll's (1979) theory of CSP and its development by Wartick and Cochran (1985), other researchers have attempted to complement the idea of CSP (Wood, 1991; Swanson, 1995; Stanwick and Stanwick, 1998). However, this thesis adheres to the original theory of Carroll (1979) for two reasons: First, competing theories are mostly only minor variations, which have not been very well established in the literature. Second, Carroll (1991) himself further developed the theory by elaborating in more detail what is meant by the categories of CSR and the stakeholder approach. For instance, economic responsibilities, such as being as profitable as possible or maintaining a high level of operating efficiency, were the primary motive of business (Carroll, 1991). Ethical and philanthropic responsibilities include being a good corporate citizen who performs and behaves in accordance with societal and ethical norms, contributes resources, and improves the quality of life (Carroll, 1991). Carroll (1991) made explicit that it is the corporate social responsibility of business to fulfill all these responsibilities.

Some people might argue that conceptual research on CSP has further developed during the last couple of decades. While this cannot be ruled out entirely, a clear majority of academic research—starting in the 1990s—focused on *empirical*[6] issues around CSP. One of the most widespread issues is the link between CSP and corporate financial performance (CFP), which I will describe later in more detail. In general, most of what follows this section refers mainly to quantitative *empirical* research, motivated by the fact that the papers constituting this thesis also exclusively deal with quantitative *empirical* issues. However, before I discuss some of the related empirical research, I will state the main terminology of this thesis.

[6]Throughout the thesis, *italic* formatting indicates an emphasis by the author.

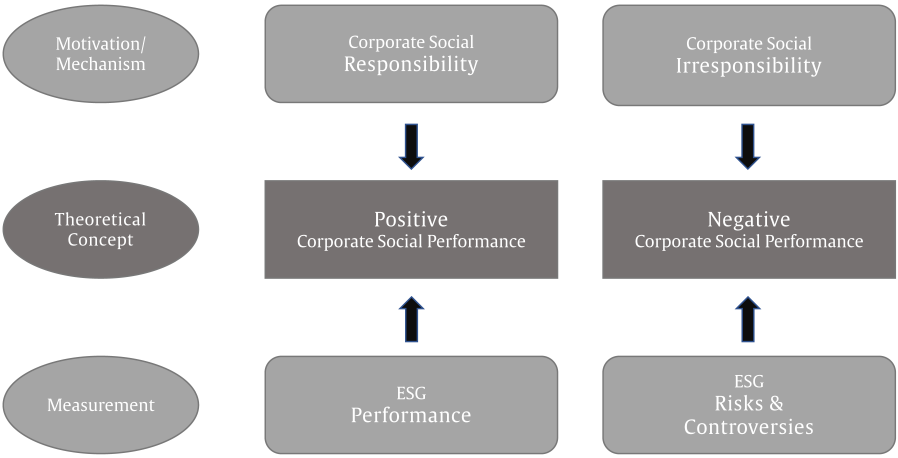
I.2 Main Terminology

This chapter clarifies the main terminology used throughout the course of this thesis. This thesis ties in with recent studies on CSR and CSP and therefore uses similar language. However, many previous studies mix positive and negative ESG issues to describe aggregate CSP. In contrast, this thesis explicitly distinguishes between positive *and* negative CSP. On the one hand, the notion of corporate social responsibility—which is defined in detail above—motivates the theoretical concept of *positive* CSP. Positive CSP captures a firm's positive engagement in ESG activities and is, thus, measured by ESG performance metrics.

On the other hand, this thesis also covers the theoretical concept of *negative* CSP. *Negative* CSP is based on the notion of corporate social irresponsibility (CSI). CSI is the counterpart of CSR and can be defined as a "set of corporate actions that negatively affects an identifiable social stakeholder's legitimate claims" (Strike et al., 2006, p. 852). CSI comprises behavior that is irresponsible, while the absence of this behavior is not necessarily responsible (Strike et al., 2006). For example, while the violation of human rights is generally perceived as irresponsible, *not* violating human rights, however, should be a matter of course. This observation also explains why corporate socially responsible behavior is hardly reported in the news, whereas socially irresponsible behavior receives high media coverage (Kölbel et al., 2017). Therefore, this study proxies for negative CSP by using ESG risks and controversies that are primarily reported by the media.

The operationalization of ESG performance and ESG risks is outlined in more detail in the research design (Chapter I.5). The differentiation between positive and negative CSP is crucial since it likely explains some of the inconclusive results of prior research (Strike et al., 2006). Figure 2 illustrates the essential terminology used throughout the course of this thesis.

Figure 2: Main Terminology



I.3 Review of Selected Empirical Literature

The empirical literature on CSP has revealed various exciting insights during the last couple of decades.[7] More specifically, this literature has been primarily but not exclusively concerned with the relationship between CSP and CFP or firm value.[8] Most of the early research found evidence that CSP is positively linked to CFP (e.g., Klassen and McLaughlin, 1996; Russo and Fouts, 1997; Waddock and Graves, 1997). This finding was also supported by more recent studies (e.g., Lev et al., 2010; El Ghoul et al., 2011; Matsumura et al., 2014; Flammer, 2015; Harjoto and Laksmana, 2018; Gupta, 2018; Flammer et al., 2019). Early evidence indicated that "it pays to be green" by finding that a higher environmental performance leads to improved stock market performance (Klassen and McLaughlin, 1996) and profitability (Russo and Fouts, 1997). An attempt to explain the positive relationship between CSP and CFP was the slack resources theory. According to this theory, it is not CSP that drives CFP but the other way

[7] This literature review only discusses quantitative empirical research on the relationship between CSP and CFP. Conceptual and theoretical research on CSR and CSP has already been addressed in Chapter I.1.

[8] Academic research uses different terms to describe corporate environmental, social, and governance responsibility and performance. The difference in meaning of these terms is mostly only marginal and research often uses them interchangeably (e.g., corporate social responsibility and corporate sustainability). However, at least slight differences exist between some terms. Therefore, this thesis consistently adheres to the previously defined term CSP and terminology as outlined in Chapters I.1 and I.2.

around (i.e., reverse causality). Indeed, the slack resources theory gained early support. High *past* financial performance (i.e., freeing up resources) was found to drive CSP (Waddock and Graves, 1997). However, the same research found that *future* financial performance is driven by CSP (Waddock and Graves, 1997).

In light of these results, McWilliams and Siegel (2000) raised criticism that the specification of the model was wrong. More specifically, they claimed that models measuring the impact of CSP on CFP should include investments in research & development (R&D) and advertising intensity since they would otherwise suffer from omitted variables bias (McWilliams and Siegel, 2000). After controlling for these variables, the positive relationship between CSP and CFP was canceled out. However, this result could not be confirmed in subsequent studies. In contrast, a massive meta-study that examined 33,878 observations from 52 research papers found robust evidence for a positive relationship between CSP and CFP (Orlitzky et al., 2003). While the social dimension of CSP seems to impact CFP strongly, the environmental dimension only does to a lesser extent (Orlitzky et al., 2003). Moreover, the relationship is stronger when using accounting-based measures of CFP compared to market-based metrics (Orlitzky et al., 2003).

Building on these results, research has been increasingly interested in the drivers of the CSP–CFP relationship. The potential positive and negative drivers of the CSP–CFP relationship are as numerous as they are diverse. They include reputation (Schnietz and Epstein, 2005; Godfrey et al., 2009), idiosyncratic risk (Lee and Faff, 2009; Sassen et al., 2016), revenue and customer satisfaction (Lev et al., 2010), environmental engagement (Makni et al., 2009), employee well-being and satisfaction (Verwijmeren and Derwall, 2010; Edmans, 2011), cost of equity capital (El Ghoul et al., 2011; Feng et al., 2015; Breuer et al., 2018; Gupta, 2018), cost of debt capital (Chava, 2014; Oikonomou et al., 2014; Goss and Roberts, 2011), stock price crash risk (Kim et al., 2014; Utz, 2018), announcement returns (Flammer, 2015), managerial entrenchment (Ferrell et al., 2016) and risk-taking (Harjoto and Laksmana, 2018), long-term orientation (Flammer et al., 2019), and legal fines (Hong et al., 2019). Most of these studies found evidence supporting a positive CSP-CFP relationship, despite the diversity of drivers and measures.

The following aims at exploring the findings of some of these studies in more detail.[9] Schnietz and Epstein (2005) found that a good reputation for CSR protects stock prices during a crisis by drawing on investor reactions to an external shock, namely, the failure of the 1999 Seattle World Trade Organization negotiations. They argued that this protection stems from a "reservoir of goodwill" of firms that are known for their environmental and social responsibility and that provides financial value in times of crisis. Godfrey (2005) developed a similar argument that CSR generates moral capital (or an insurance-like effect) that protects shareholder value. In a more generalizable research setting than Schnietz and Epstein (2005), Godfrey et al. (2009) found evidence supporting the notion of an insurance-like effect of institutional CSR (i.e., activities that aim at secondary stakeholders or the whole society). However, they did not find such an effect for technical CSR (i.e., activities targeting specific trading partners). Two studies found that employee well-being and satisfaction is positively related to shareholder returns and negatively associated with bankruptcy risk (Verwijmeren and Derwall, 2010; Edmans, 2011).

Another stream of literature has focused on the environmental and social dimensions of CSP. One study suggested that the environmental dimension negatively impacts CFP (Makni et al., 2009). However, the majority of more recent studies found consistent evidence that environmental performance is positively related to financial performance and firm value (Chava, 2014; Matsumura et al., 2014; Jo et al., 2015). The same applies to the value-enhancing role of the social dimension generally and of philanthropic engagement particularly, which have received considerable empirical support (Orlitzky et al., 2003; Lev et al., 2010; Goss and Roberts, 2011).

Another strand of research has examined the cost of equity and debt capital as a potential value-enhancing or -destroying driver in the CSP–CFP relationship. The majority of research found evidence for reduced cost of equity and debt capital due to CSR and CSP (El Ghoul et al., 2011; Oikonomou et al., 2014; Feng et al., 2015; Breuer et al., 2018; Gupta, 2018). These results are supported by research finding that CSP decreases risk on various dimensions and therefore lowers cost of equity and debt capital. These risk dimensions include idiosyncratic

[9] For brevity, I only discuss a selection of relevant studies in a broader context. Each paper of this dissertation (cf., Chapters II, III, and IV), as well as the derivation of the research ideas (cf., Chapter I.4), have its standalone discussion of relevant literature.

risk (Lee and Faff, 2009; Sassen et al., 2016), credit risk (Kölbel et al., 2017), stock price crash risk (Kim et al., 2014; Utz, 2018), and the risk-taking behavior of managers (Harjoto and Laksmana, 2018).

Moreover, research found that CSR is connected to more positive announcement returns (Flammer, 2015), less managerial entrenchment (Ferrell et al., 2016), higher market value (Aouadi and Marsat, 2018), lower sanctions in corruption-related litigation cases (Hong et al., 2019), and more long-term orientation by executives (Flammer et al., 2019). In general, these factors seem to have a positive impact on CFP in one way or another.

Table 1 summarizes the results of selected studies. However, the synopsis and the remarks made in this chapter are not exhaustive and just highlight studies that are related to this thesis' research objective. Please refer to meta-studies and literature reviews, such as Orlitzky et al. (2003), Margolis et al. (2009), and Grewatsch and Kleindienst (2017), for a more comprehensive overview of the large body of literature on CSP and CFP.

I have briefly outlined these studies because they open interesting areas of future research comprised of many potentially relevant questions about the main drivers of the CSP–CFP relationship that have not yet been answered. Moreover, questions as to what investors and stakeholders can learn from this relationship remain open. This thesis aims at answering some of these questions that can be derived from prior research. This prior research primarily includes materiality research by Khan et al. (2016), research on risk management implications of CSP (Cai et al., 2016; Godfrey, 2005; Godfrey et al., 2009), and research on corporate social irresponsibility and credit risk by Kölbel et al. (2017). The next chapter identifies and outlines the research ideas on which the thesis at hand focuses on.

Table 1: Selected Literature on Corporate Social Responsibility and Performance

Author & Year	Sample	Key Findings	CSP-CFP Link
Klassen and McLaughlin (1996)	N=110	Strong (weak) environmental management leads to significantly positive (negative) returns	Positive

Table 1 continued from previous page

Russo and Fouts (1997)	N=486	A higher environmental performance leads to stronger economic performance, with high-growth industries positively moderating this relationship	Positive
Waddock and Graves (1997)	N=486	CSP positively influences financial performance	Positive
McWilliams and Siegel (2000)	N=524	Including R&D investment into the model cancels out the positive effect of CSR	Neutral
Orlitzky et al. (2003)	Meta-Study	Both environmental and social responsibility pay off financially, but social does more strongly.	Positive
Schnietz and Epstein (2005)	N=416	A good reputation for social responsibility protects firm value during a crisis	Positive
Godfrey et al. (2009)	N=178	Institutional CSR provides insurance-like effects while technical CSR does not	Positive/ Neutral
Lee and Faff (2009)	N=11,479	CSP leads to significantly lower idiosyncratic risk and this risk may be priced by the market	Positive
Makni et al. (2009)	N=179	Aggregate CSP has no effect on CFP while the environmental dimension negatively impacts CFP	Neutral/ Negative
Lev et al. (2010)	N=1,618	Charitable contributions are significantly associated with future revenue	Positive
Verwijmeren and Derwall (2010)	N=7,494	Higher employee well-being significantly reduces the probability of bankruptcy (due to lower debt ratios)	Positive
Edmans (2011)	N=69	Firms with high employee satisfaction earn abnormal returns and exhibit significantly more positive earnings surprises and announcement returns	Positive
El Ghoul et al. (2011)	N=12,915	CSR reduces the cost of equity capital	Positive

Table 1 continued from previous page

Goss and Roberts (2011)	N=3,996	Fewer social responsibility concerns lead to significantly lower cost of bank loans	Positive
Chava (2014)	N=13,114	Fewer environmental concerns lead to significantly lower cost of debt	Positive
Kim et al. (2014)	N=12,978	CSR decreases stock price crash risk and is thus value-enhancing	Positive
Matsumura et al. (2014)	N=1,365	Lower carbon emissions lead to higher firm value	Positive
Oikonomou et al. (2014)	N=7,794	High CSP leads to lower bond yield spreads (i.e., lower costs of debt)	Positive
Feng et al. (2015)	N=10,803	High CSR is significantly associated with reduced cost of equity capital	Positive
Flammer (2015)	N=2,729	CSR proposals lead to positive announcement returns and superior accounting performance	Positive
Jo et al. (2015)	N=4,895	Corporate environmental responsibility enhances return on assets with a time-lag of 1 or 2 years in the financial services sector in Europe but not in Asia	Positive/ Neutral
Ferrell et al. (2016)	N=97,373 N=7,048	CSR positively drives firm value and mitigates the negative relation between managerial entrenchment and firm value	Positive
Khan et al. (2016)	N=14,388	Material sustainability issues positively drive CFP. Immaterial sustainability issues do *not* drive CFP	Positive/ Neutral
Sassen et al. (2016)	N=8,752	CSP increases firm value through lower idiosyncratic risk	Positive
Kölbel et al. (2017)	N=9,939	Less CSI reduces credit risk and thus increases firm value	Positive
Aouadi and Marsat (2018)	N=8,321	CSP positively affects market value but only for high-attention firms	Positive
Breuer et al. (2018)	N=19,183	The impact of CSR on the cost of equity capital depends on country-specific investor protection	Neutral

Table 1 continued from previous page

Gupta (2018)	N=23,301	Improvements in environmental practices lead to a reduced implied cost of equity	Positive
Harjoto and Laksmana (2018)	N=10,153	CSR positively influences firm value by reducing excessive risk-taking	Positive
Utz (2018)	N=1,339 N=7,433 N=3,758 N=9,014	CSR mitigates crash risk in Europe and the United States (US) and is thus value-enhancing	Positive
Flammer et al. (2019)	N=4,533	Integrating CSR criteria in executive compensation increases long-term orientation and firm value	Positive
Hong et al. (2019)	N=131	Socially responsible firms receive 65% lower sanctions in corruption-related litigation cases	Positive

Notes: This table shows selective literature, which is relevant to the thesis' research objective in a broader sense. Each paper of this thesis has its own in-depth discussion of closely related literature. Please also note that this synopsis is not exhaustive and displays only a fraction of the analyzed literature. The selection of studies was deemed most relevant to the thesis' objective by the author. Studies that appeared in top-tier journals were given priority (i.e., ranked A or higher). If more than one *N* is specified, the respective study used more than one main sample. The studies were primarily sorted chronologically to show the evolution in CSP research. On a secondary level, the studies were sorted alphabetically.

I.4 Research Ideas and Objectives

I identified three research ideas in the context of positive and negative CSP, which have been neglected by previous research. The research ideas are primarily linked to materiality research (Khan et al., 2016), research on risk management implications of CSP (Cai et al., 2016; Godfrey, 2005; Godfrey et al., 2009), and research on corporate social irresponsibility and credit risk (Kölbel et al., 2017).

I.4.1 Corporate Social Performance, Earnings Quality, and Materiality

The first research idea addresses the topical issue of materiality in an accounting and CSP context. It investigates to which extent information on material and non-material CSP is linked to earnings quality (EQ). I build on materiality research by Khan et al. (2016), the extensive em-

pirical literature on EQ, and two fundamental streams of CSP theories: ethical and reputational theories, and self-interest and instrumental theories. The materiality of ESG issues has long been neglected by academic research. Subsequently, research perceived ESG issues as equally relevant to all industries and thus almost no methodological differences between industries have been made. Khan et al. (2016) were the first to systematically address the materiality of ESG issues and their impact on CFP. Their main finding was that only the material ESG issues positively drive CFP and not the non-material ones. This finding implicated that it is essential for investors and analysts to exclusively focus on the material ESG issues for each industry to determine the impact of CSP on CFP.

While Khan et al. (2016) found that it is crucial to only consider material ESG issues on a meta-level, many open questions about the drivers of this relationship remained unexplored. For example, one of these potential drivers is the earnings quality of a firm. EQ is relevant from an investor's and analyst's perspective since it is an essential factor in equity valuations. Even though previous studies have looked at the relationship between CSP and EQ, none of these studies considered the materiality of ESG issues (e.g., Chih et al., 2008; Martínez-Ferrero et al., 2016; Hong and Andersen, 2011; Kim et al., 2012). This lack of consideration is striking since the differentiation between material and non-material CSP potentially plays a vital role in explaining the mixed results obtained by previous research. Therefore, this paper introduces the materiality dimension to the relationship between CSP and EQ. It investigates whether differentiating between material and non-material CSP explains EQ.

On the one hand, managers that care about the ethical and reputational implications of their CSP should be more likely to be interested in performing well across those ESG issues that are material (i.e., value-relevant). On the other hand, managers that follow their self-interest (or an instrumental motive) are more likely to use CSP as a disguise ("greenwashing") from their earnings management. This paper states that following such a strategy of disguise is only possible when firms are trying to perform well across the numerous non-material ESG issues (or simply as many as possible) as compared to the few material ones. The two theory streams (ethical and reputational theories, and self-interest and instrumental theories) are therefore the foundation of the first paper. Both the existing literature and the theories, as well as the relevant hypotheses, will be discussed later in more detail under Chapter II. While the first paper focuses

on the accounting perspective, the second research idea additionally introduces the market perspective.

I.4.2 Corporate Social Performance and Litigation Risk

The second research idea concerns the impact of CSP on risk and market value in a litigation context. Litigation can heavily impact firms on many dimensions, including increased CEO turnover (Aharony et al., 2015), less use of external debt and equity financing (Autore et al., 2014), and significantly negative stock price reactions (Gande and Lewis, 2009). Therefore, litigation risk (i.e., the probability of facing a class action lawsuit) is highly relevant to investors and analysts for making investment decisions and recommendations. Research has also shown that CSP offers risk-reducing characteristics (Cai et al., 2016; Harjoto and Laksmana, 2018; Benlemlih et al., 2018) and insurance-like effects in times of corporate crisis (Godfrey, 2005; Koh et al., 2014). Consequently, the second paper tests whether positive and negative CSP reduces litigation risk (i.e., has risk-reducing characteristics regarding class action lawsuits). It further examines whether CSP provides a moderating effect on the relationship between litigation and market value (i.e., offers insurance-like protection during class action lawsuits).

These two research objectives are based on two fundamental theories: risk management theory and stakeholder theory. The first theory, risk management theory, suggests that CSP has risk-reducing characteristics due to decreased information asymmetries (Cui et al., 2018), which translate into lower risk (Benlemlih et al., 2018), and by functioning as a risk management tool itself (Cai et al., 2016; Harjoto and Laksmana, 2018). These risk-reducing characteristics indicate that firms having a high positive CSP and a low negative CSP might also experience a lower litigation risk (i.e., fewer class action lawsuits). Therefore, the first research objective is to test whether CSP impacts the probability of a firm facing a class action lawsuit. The second theory, stakeholder theory, addresses the potential of CSP providing insurance-like effects during corporate crisis. The potential moderating effect of CSP is based on the idea of a "reservoir of goodwill" that is being generated among stakeholders by engaging in ESG issues (Koh et al., 2014; Freeman, 2010). According to this notion, CSP is a strategic investment in developing firm integrity among stakeholders (McWilliams et al., 2006), which is then likely to pay off in

a corporate crisis. Hence, the second research objective is to test whether CSP protects firm market value in cases where a class action lawsuit has already occurred.

The second research idea also synthesizes the findings on both research objectives by implementing a trading strategy. Using the paper's newly introduced model, the trading strategy focuses on those firms with the lowest predicted probability of class action lawsuits (=litigation risk) and with the highest ESG performance. The strategy is based on the idea that firms with an exceptionally low litigation risk potentially yield positive alpha due to lower costs caused by litigation (e.g., legal fees, reputational damage, lower revenue, etc.). Moreover, if such a firm still faces litigation, ESG performance is likely to provide some insurance-like protection, increasing alpha even further compared to firms that face litigation but have a low ESG performance. Therefore, combining low litigation risk with high ESG performance in a trading strategy is an attempt to generate positive alpha for investors.

I.4.3 Corporate Social Performance and Credit Risk Prediction

The third paper addresses another common challenge faced by investors and analysts: the prediction of corporate credit risk. Hence, the research idea is concerned with improving credit risk prediction by using advanced machine learning (ML) algorithms and data on negative CSP (i.e., ESG risks). Previous research has shown that ML techniques can provide superior explanatory power, especially in financial prediction tasks such as forecasts of consumer credit risk, security pricing, portfolio construction, financial time series prediction, and the prediction of sovereign credit ratings (Khandani et al., 2010; Niu et al., 2020; Bennell et al., 2006; Gu et al., 2020). Moreover, in the context of default prediction, the prediction accuracy of ML techniques recently gained considerable support (Barboza et al., 2017). The study of Barboza et al. (2017) showed that advanced ML algorithms provide significantly higher prediction accuracy in predicting default, which is a common classification problem (1 or 0). The third paper of this thesis tests whether the superior explanatory power also holds for credit risk prediction in a regression design (as compared to default risk and a classification design).

In addition to using an ML regression technique, the paper uses data on ESG risks in a first attempt to improve credit risk prediction. It also intends to test whether ESG risks have

been priced in by rating agencies in their credit rating process during the period 2007 to 2019. Major rating agencies have repeatedly claimed that ESG considerations play a role in their credit rating process.[10] The integration of ESG risks into the analysis is also motivated by an investor survey conducted by the Principles for Responsible Investments (PRI) association. The survey found that—among other things—the majority of credit analysts, portfolio managers, and ESG analysts deemed ESG issues as relevant in their organizations' valuation and credit risk models.

In general, previous research has repeatedly shown that CSP reduces default risk (e.g., Sun and Cui, 2014) and firm risk (e.g., El Ghoul et al., 2011; Lee and Faff, 2009). More specifically, socially and environmentally responsible firms experience higher credit ratings (Hsu and Chen, 2015; Jiraporn et al., 2014; Kiesel and Lücke, 2019) and lower bond risk premia (Goss and Roberts, 2011) as well as significantly lower interest rates on bank loans (Chava, 2014). Positive CSP has been further linked to *lower* credit default swap (CDS) spreads (Drago et al., 2019), indicating lower credit risk. Research also found that negative CSP (i.e., ESG risks) significantly *increases* CDS spreads (Kölbel et al., 2017). Concluding, these results reasonably suggest that ESG risks impact credit risk and should thus be priced in the credit ratings of major rating agencies. Therefore, the third research objective is to examine whether using ML algorithms and data on ESG risks can improve the prediction accuracy of credit risk models.

I.4.4 Similarities and Differences Between the Papers

Regardless of the different research objectives, the first and second papers share important similarities, especially regarding the main concepts as well as the implications for investors and analysts. More specifically, among the similarities between the first and second papers are the ESG data from Thomson Reuters, the conclusion that information on CSP is decision-relevant, and the implication that investors and analysts can benefit from considering the CSP of firms. They can partly benefit from information on CSP due to a better assessment of the earnings quality of companies as well as more informed decisions about the litigation risk and potential

[10] For example, see Standard & Poor's (S&P) Global's claims available via: https://www.spglobal.com/ratings/en/products-benefits/products/sustainable-finance and https://www.spglobal.com/ratings/en/products-benefits/products/esg-in-credit-ratings#. Fitch issued a similar claim but later than S&P Global: https://www.fitchratings.com/site/pr/10058528.

costs of litigation. However, while the first paper supports the positive CSP–CFP link only to a limited extent, the second paper provides clear evidence that a higher positive CSP and a lower negative CSP positively contribute to the CFP and firm value. Both papers make use of extensive ESG performance and controversies data from Thomson Reuters. However, while the first paper focuses on Europe, the second paper investigates US firms solely due to the US-specific phenomenon of class action lawsuits. Moreover, on the one hand, the first paper focuses on the materiality of ESG performance (i.e., positive CSP). On the other hand, the second paper includes both ESG performance (i.e., positive CSP) and ESG controversies (i.e., negative CSP).

Similar to the first two papers, the third paper builds on the value-relevance of CSP to investors and analysts by examining credit risk. However, instead of following an approach of statistical inference (cf., Papers 1 and 2), the third paper focuses on a predictive approach using trained machine learning models. Therefore, the paper is mainly concerned with the final out-of-sample prediction and accuracy of the models and does not allow further conclusions about the underlying mechanisms (economical and theoretical) between the predictor and target variables. The third paper also nicely complements the first and second papers in terms of data. It uses ESG risk assessments from a "new generation" business intelligence provider, RepRisk, whose proprietary index is based on third-party data. In contrast, the first and the second papers use ESG ratings from Thomson Reuters, which, while highly accepted in academic research, are mainly based on self-reported data by firms. In terms of methodology, the third paper clearly shows that using more advanced methods, which are capable of capturing non-linearity in the data, yields superior results—at least in a prediction context.

All three papers jointly refer to corporate social performance as the most widespread framework to capture corporate environmental, social, and governance activities. On the one hand, they build on previous studies showing that CSP is relevant to the CFP in various ways. On the other, each paper identifies potential and novel drivers of the relationship between CSP and CFP—direct and indirect drivers. While the first paper investigates EQ, an important figure in investors' equity valuations, the second paper examines actual market impacts of positive and negative CSP directly. The third paper tests whether negative CSP is priced in credit ratings, which in turn affect the cost of capital (i.e., firm value). After all, the papers evolve from an accounting-based perspective of CSP (cf., Paper 1) to a market-based perspective (cf., Paper

2) and, finally, a predictive approach (cf., Paper 3). They share important implications for investors, analysts, and corporations as to how environmental, social, and governance activities influence financial performance and firm value. Table 2 provides an overview and comparison of the three papers constituting the main body of this thesis.

Table 2: Thesis Overview

Paper	Research Question(s)	Objective/Hypotheses	Depend. Var	Ind. Var.	Main Findings	Implications
1st	How does material and non-material CSP influence earnings quality?	*Transparency Hypothesis* $H1_A$: Firms with high CSP exhibit more earnings quality than firms with low CSP *Instrumental Hypothesis* $H1_B$: Firms with high CSP exhibit less earnings quality than firms with low CSP *Materiality Hypothesis* $H2_A$: Material CSP firms exhibit more earnings quality than non-material CSP firms *Immateriality Hypothesis* $H2_B$: Non-material CSP firms exhibit less earnings quality than material CSP firms	Accruals Quality Real Earnings Management	CSP Material CSP Non-Material CSP	Material CSP firms exhibit higher accruals quality (i.e., higher earnings quality) Material and non-material CSP firms exhibit more real earnings management (i.e., lower earnings quality) After controlling for industry regulation, the difference in accruals quality due to material CSP does not hold The higher earnings quality (measured by accruals quality) is driven by highly regulated sample firms	Material CSP is not necessarily a useful indicator of higher earnings quality Firms use CSP to disguise from real earnings management Accruals quality and real earnings management are substitutes

Table 2 continued from previous page

Paper	Research Questions	Objective/Hypotheses	Depend. Var.	Ind. Var.	Main Findings	Implications
2nd	How does positive and negative CSP influence the probability of class action lawsuits? To what extent does positive CSP protect market value around a class action lawsuit filing?	$H1_A$: Higher positive CSP (i.e., a stronger ESG performance) is linked to a lower current-year probability of facing a class action lawsuit $H1_B$: Lower negative CSP (i.e., fewer previous-year ESG controversies) is linked to a lower current-year probability of facing a class action lawsuit $H2$: Firms with higher positive CSP (i.e., stronger ESG performance) exhibit smaller losses in terms of (cumulative) abnormal returns during class action lawsuits than firms with lower positive CSP	Prob (Litigation=1) $CAR_{t=10}$	ESG_P-score ESG_N-score	A one standard deviation improvement in ESG controversies reduces the predicted probability of a class action lawsuit by 22.58% Firms with a high positive CSP experience only half as high negative abnormal returns around a class action lawsuit filing compared to firms with low or mediocre CSP	Scores on the mere number of ESG controversies provide useful information about litigation risk to stakeholders Lower litigation risk due to fewer ESG controversies is likely to translate into lower equity risk premia Positive CSP mitigates financial constraints and potentially prevents firms from financial distress during litigation

Table 2 continued from previous page

Paper	Research Questions	Objective/Hypotheses	Depend. Var.	Ind. Var.	Main Findings	Implications
3*rd*	Do more advanced machine learning algorithms provide superior predictive ability for credit risk? Is negative CSP adequately reflected in credit risk as measured by credit ratings?	More advanced machine learning algorithms should be superior to traditional approaches due to the ability to adequately capture non-linear and complex relationships in the data ESG issues are supposed to play a major role in firm risk and credit risk. Rating agencies claim that they include ESG considerations in their ratings. Thus, ESG risks should be informative to the prediction of credit risk	Long-term issuer credit ratings from S&P, Moody's, and Fitch	Current RRI Peak RRI Vector of ESG Issues	Advanced machine learning algorithms clearly excel traditional approaches in the predictive ability for credit risk ESG risks are not (yet) adequately reflected in credit risk as measured by long-term issuer credit ratings	Advanced machine learning algorithms are capable of capturing the non-linear and complex relationships within the data to improve credit risk prediction significantly. Investors and analysts should consider using algorithms such as Random Forest regression when predicting credit risk Even though previous research suggests that ESG risks play a major role in firm performance and risk, ESG risks are not (yet) systematically considered in credit ratings. Given the relevance of ESG risks, rating agencies are likely to systematically consider ESG risks in their rating process and decisions in the future

Notes: This table provides an overview of the three papers that constitute the main body of this thesis. $CAR_{t=10}$ denotes the cumulative abnormal return on day 10 after the class action lawsuit filing. Prob(Litigation=1) denotes the probability of facing litigation as measured by one of the litigation variables. All other terms have already been described. More information on the variables and hypotheses can also be found under the respective chapters of the papers.

I.5 Overview of the Data and the Research Design

This thesis leverages three distinct data sets and a wide-ranging methodology, including multiple regression (linear and Probit), event study methodology, and a comprehensive set of machine learning algorithms, to investigate the above-mentioned research objectives. All three papers make use of partially distinct but ultimately *quantitative empirical* methods. Papers one and two use methods of statistical inference (i.e., in-sample), and paper three follows a predictive approach using machine learning (i.e., out-of-sample). While the first paper exploits the advantages of statistical inference from robust multiple regression, the second paper additionally adapts event study methodology and a propensity score matching approach to account for potential reverse causality and omitted variables bias.[11] Moreover, the second paper applies a portfolio approach to test the investment management implications of the previously obtained results. The third paper progresses methodologically even further by moving from statistical inference to making out-of-sample predictions with trained machine learning models.

This thesis uses a minimum significance level of 5% as this is the universally recognized standard in finance and accounting. Results with p-values exceeding 5% are, thus, considered statistically insignificant throughout the course of this thesis. After all, the three papers are methodologically quite distinguished from each other and embody the "evolutionary" methodological development of this thesis. Therefore, I will separately explain the data sets and methods used in each of the three papers in the following.

I.5.1 1st Paper: European Data and Multiple Linear Panel Regression

This paper uses a European sample of 542 firms from 27 different industries in the period from 2005 to 2016. The main reason for using European data is that previous research has only used US data so far. Thus, this thesis closes the geography-related research gap. Using raw ESG data from Thomson Reuters, the paper particularly stands out for recalculating ESG scores based on the materiality guidance of the Sustainability Accounting Standards Board (SASB).

[11] All papers of this thesis use lagged independent variables to provide some baseline robustness of the results against reverse causality. However, previous research argued that event study methodology and propensity score matching even further reduces the risk that the results suffer under reverse causality and omitted variables bias (e.g., Hawn et al., 2018).

Hence, it follows the Thomson Reuters benchmarking procedure of ESG scores and introduces the materiality of ESG issues at the same time (i.e., distinguishing between material and non-material issues). All 406 Thomson Reuters ESG issues were classified by hand as material or non-material for 27 industries separately.

Performing such a recalculation procedure of Thomson Reuters ESG scores has been unprecedented in academic literature. However, Khan et al. (2016) applied a similar approach to data from Kinder, Lydenburg, and Domini (KLD) in a different context. Most of the previous research relied on data from KLD, whereas this paper leveraged the advantages of Thomson Reuters ESG data, which I describe later under Chapter II. In addition to the ESG data, I used market- and accounting-related data from Compustat and The Center for Research in Security Prices (CRSP) provided through Wharton Research Data Services (WRDS). Control variables include, for example, working capital, firm age, market-to-book ratio, return on assets (ROA), and leverage. The regression approach mainly consists of in-sample multiple panel regressions using ordinary least squares (OLS). More details on the exact recalculation procedure, the dependent and independent variables, and sample characteristics can be found in Chapter II.

I.5.2 2nd Paper: US Data, Multiple Probit Panel Regression, Event Study Methodology, and Portfolio Approach

The main research object of the second paper is US class action lawsuits. Therefore, this paper uses open-source class action data from the Securities Class Action Clearinghouse (SCAC) provided by Stanford Law School and Cornerstone Research. The class action data goes back until 1996 and is updated daily (business days).[12] However, the period of investigation is from 2003 to 2017 since ESG data is not available on a larger scale before 2003. Data on class action periods and class action filings of US firms are collected back until the year 2000 to construct the class action (i.e., litigation risk) variables. The final sample comprises 734 class period years and 303 actual lawsuit filings (from a total of 7571 firm-year observations).

In addition to the class action data, this paper makes use of ESG performance and ESG

[12] The data is available via the SCAC website: http://securities.stanford.edu/index.html

controversies data from Thomson Reuters. While the use of ESG performance data is similar to the first paper, the use of data on the involvement in ESG controversies of firms is an additional dimension. The ESG controversy measure of Thomson Reuters captures the number of controversies a firm experienced during the previous year and benchmarks this number against the firm's industry peers. Thomson Reuters deems the measure as a viable standalone indicator of ESG controversies (or negative CSP).

In a first step, the paper makes use of in-sample multiple Probit panel regression to determine the impact of ESG performance and ESG controversies on the risk of facing a class action lawsuit (i.e., litigation risk). In a second step, it uses state-of-the-art event study methodology and a propensity score matching approach to determine the market impact of ESG performance around the filing of a class action lawsuit. Therefore, the event study is based on a market-based estimation of the expected returns as well as on a control firm's return, which was determined by propensity score matching.

In a third and final step, the study uses a portfolio approach to test whether a trading strategy that is based on the paper's previous results can yield positive alpha. More specifically, the portfolios are built according to predictions of litigation risk and the ESG performance of a firm. The data from 2003 to 2012 is used to estimate the Probit model's parameters. The data from 2013 to 2017 is used to build the portfolios. Using cutoffs of 10%, the various portfolios contain firms that have high litigation risk, low litigation risk, high ESG performance, and low ESG performance. The trading strategy also comprises a portfolio that includes firms with the 10%-lowest litigation risk and the 10%-highest ESG performance. The portfolios are tested for alpha by using a Fama-French 3-factor model and a Fama-French 5-factor model. More details on the class action and ESG data, as well as the methodology (i.e., multiple panel Probit regression, event study, and portfolio approach), can be found in Chapter III.

I.5.3 3rd Paper: Global Data and Machine Learning Regression Design

The third paper focuses on out-of-sample credit risk prediction using ML algorithms that are capable of capturing non-linear and complex relationships in the data to improve predictions. The paper stands out in particular for the application of a wide range of 12 ML algorithms. The

algorithms are split into two categories: so-called "traditional" and "advanced" algorithms. The paper assigns conventional methods from statistics (including regularization) to the "traditional" category; namely, Linear regression, ElasticNet, Ridge, OrdinalRidge, Support Vector Regression (SVR) with a linear kernel. All the other, mostly non-parametric methods are assigned to the "advanced" category; namely, SVR with a radial basis function (RBF) kernel, Decision Tree regression (DTR), Random Forest regression (RFR), AdaBoost regression (AdaBoostR), Gradient Boosting regression (GBR), K-nearest Neighbors regression (KNNR), and Multi-layer Perceptron regression (MLPR). Only a couple of programming languages offer such a comprehensive and sound repertoire of algorithm implementations (e.g., R and Python). The paper at hand uses Python 3.7.4 and its affiliated ML library *sklearn*[13] to implement the algorithms. *sklearn* is an open-source application programming interface (API) that provides simple and efficient tools for predictive data analysis using machine learning. Above that, the scientific integrated development environment (IDE) for Python, Spyder, which is also open-source, is used for coding and implementation.

In addition to the ML part, the paper exploits data on ESG risks from RepRisk. RepRisk provides business intelligence by specializing in ESG and business conduct risk research using the power of artificial intelligence and big data to derive relevant risk metrics. The research process of RepRisk will be explained later under the respective section of the paper. It is, however, more important to note that the RepRisk measure of ESG risk substantially differs from the ratings on ESG performance and controversies from Thomson Reuters used in the first and the second paper. While Thomson Reuters relies to a large extent on self-reported, non-financial information of firms, RepRisk uses more objective information from independent news media sources to derive their metrics. Therefore, RepRisk belongs to a new generation of business intelligence providers that actively integrate independent media on a large scale into their rating process.

Moreover, RepRisk focuses on capturing ESG risks (i.e., negative CSP) and not ESG performance (i.e., positive CSP) due to the immediate and well-founded relevance of ESG risks for the financial and operating performance of firms. The third paper makes use of this data from

[13] *sklearn* is the short version for "scikit learn". The application programming interface (API) reference can be found on https://scikit-learn.org/stable/index.html

RepRisk. More specifically, it uses the proprietary and state-of-the-art risk metric, the RepRisk Index (RRI), to measure the exposure of firms to environmental, social, and governance risks. Table 3 provides an overview and compares the data and methodology used in the course of this thesis.

Table 3: Overview of the Data, Methodology, and Test Statistics

Paper	Data Sources	Geography	Sample(s)	Methodology	Test Type(s)
First	Thomson Reuters ESG, Sustainability Accounting Standards Board, Compustat, The Center for Research in Security Prices	Europe	Panel data, N=5974, 542 firms, 27 industries	Multiple panel regression, Country FE, Industry FE, Two-way adjusted standard errors	Parametric
Second	Securities Class Action Clearinghouse (Stanford Law School), Thomson Reuters ESG, Compustat, The Center for Research in Security Prices	US	Panel data, N=7671, 1796 firms, all industries	Multiple Probit panel regression, Year FE, Two-way adjusted standard errors, Event study, Propensity score matching, Portfolio approach	Parametric
Third	RepRisk, Compustat, Capital IQ, Thomson Reuters	Global	N=43476, training N=34780, test N=8696, all industries,	Linear regression, ElasticNet, Ridge, OrdinalRidge, Linear Support Vector regression, RBF Support Vector regression, Decision Tree regression, Random Forest regression, AdaBoost regression, Gradient Boosting regression, K-nearest Neighbors regression, Multi-layer Perceptron regression	Parametric, Non-Parametric

Notes: This table provides an overview of the relevant data and methodology that have been used in this thesis. More details on the data and methodology can be found under the respective sections in the papers. Some of the methodology used in robustness tests is not outlined in this table since it mostly differs only slightly from the methodology in the main part (e.g., the algorithms of the classification design used alternatively in the third paper).

I.6 Structure

The main part of this thesis is structured in three chapters (i.e., one for each paper) and a concluding chapter. The first paper (Chapter II) investigates the impact of material and non-material CSP on the earnings quality of a European sample of 542 firms from 27 different industries. After having analyzed the potential relevance of material ESG issues for earnings quality and how materiality can be integrated into ESG research, the second paper (Chapter III) moves from the accounting-based impact of CSP to a market-based approach. Chapter III analyzes risk-mitigating and insurance-like effects of CSP in the context of US class action lawsuits. The third paper (Chapter IV) moves from a market-based impact assessment of CSP to a predictive approach. It investigates the potential of using information on ESG risks and an extensive machine learning pipeline to improve credit risk prediction. Chapter V closes with some conclusions on the main results and contributions, a general note on limitations, and some avenues of future research.

I.7 Data and Code Repository

In accordance with scientific conventions, the data and code used in this thesis are available in the Swiss-European non-commercial data repository called Zenodo. Zenodo is an initiative by CERN in Geneva guaranteeing at least 20 years of data conservation. Zenodo follows the FAIR guiding principles for scientific data management and stewardship (cf., Wilkinson et al., 2016). FAIR stands for findable, accessible, interoperable, and reusable. The long-term storage of the data considerably increases the replicability and traceability of the analysis and results of this thesis. For instance, it also allows to use the trained ML models from the third paper on new and unseen data. The repository comprises raw, cleaned, and processed data as well as the Python and Stata code used in this thesis. It is available via the following link: https://dx.doi.org/10.5281/zenodo.3904416. Access can be granted on request from the author.

I.8 Publication Status and Strategy

The first paper, entitled "Corporate Social Performance and Earnings Quality: The Role of Materiality," was published in the Journal of Environmental Law & Policy on June 7th, 2019. The journal is ranked as *B* in VHB-JOURQUAL3 and is listed in the American Economic Association EconLit journal list. In the Handelsblatt ranking, it has been repeatedly classified as a B-journal. The journal follows a double-blind peer-review process. The second paper has been accepted for publication at the Financial Analysts Journal. The journal is ranked as *B* in VHB-JOURQUAL3, and its impact factor was 1.816 in 2018. The third paper has been accepted for publication at the journal Credit and Capital Markets. The journal is ranked as *C* in VHB-JOURQUAL3.

II Paper 1: Corporate Social Performance and Earnings Quality: The Role of Materiality[14]

Daniel V. Fauser

Abstract This study tests whether firms with high material CSP perform better on earnings quality (EQ). I use a cross-sectional sample of 542 European firms from 27 different industries and materiality guidance from the Sustainability Accounting Standards Board (SASB).[15] I develop a new measure of material and non-material CSP by recalculating Thomson Reuters ESG scores based on the guidance from SASB. The state-of-the-art measures for EQ are: accruals quality (AQ) and real earnings management (REM). In general, results indicate that firms performing well on material ESG issues (i.e., material CSP firms) have higher accruals quality. However, I also find that both material and non-material CSP firms engage in more REM. In robustness checks, I corroborate that, in regulated industries, both material and non-material CSP firms show higher AQ. The findings indicate that material CSP is not necessarily a reliable indicator of higher EQ. However, in regulated industries, more socially and environmentally responsible firms tend to refrain from managing their earnings. Moreover, I find that firms tend to use real earnings management as a substitute for manipulating accruals.

Key Words: corporate social performance; earnings quality; materiality; ESG

[14]This paper has been presented at the conference Managing and Financing Responsible Businesses, the inaugural conference of the Global Research Alliance for Sustainable Finance and Investment, in Maastricht from September 5th to 7th, 2018, and at the ACA Symposium at the University of St.Gallen, from November 7th to 9th, 2018. On June 7th, 2019, the paper was published in the Journal of Environmental Policy & Environmental Law (Zeitschrift für Umweltpolitik & Umweltrecht). The journal is ranked as *B* in VHB-JOURQUAL3 and is listed in the American Economic Association EconLit journal list. In the Handelsblatt ranking, it has been repeatedly classified as a B-journal. The journal follows a double-blind peer-review process.

[15]The Sustainability Accounting Standards Board kindly provided the materiality data under an academic license agreement. © 2014 The SASB Foundation. All Rights Reserved.

II.1 Introduction

The awareness of CSP has grown dramatically in the last few years. The reasons for this increased awareness are numerous. Environmental developments such as rising sea temperatures have caused considerable damages to coral reefs like the Great Barrier Reef. In 2017, with Hurricanes Harvey, Irma, and Maria, the United States (US) experienced one of the worst hurricane seasons since recording began. Science mostly agrees that the business activities of multinational firms contribute to developments like climate change. Ethicists, economists, and politicians increasingly point to the responsibility of firms to internalize the environmental and social consequences of business activities. The concept of CSP is an approach to capture the degree to which firms take such a responsibility.[16] However, the credo "profit above all" still seems to dominate the world of business to a high degree.

Building on this credo, previous studies have examined whether CSP also has financial implications for firms (e.g., El Ghoul et al., 2011; Ferrell et al., 2016; Matsumura et al., 2014; Orlitzky et al., 2003; Waddock and Graves, 1997) since these are particularly relevant for investors. EQ, as one of the key figures, is of particular interest to investors. EQ is not only a question of accounting integrity, but investors and analysts use earnings in their equity valuations. If the earnings used in these equity valuations are of higher quality, investors can more reliably predict the firm's value and future cash flows. Hence, earnings of higher quality used in these valuations lead to better investment decisions. Consequently, some studies have linked CSP to EQ (e.g., Chih et al., 2008; Martínez-Ferrero et al., 2016; Hong and Andersen, 2011; Kim et al., 2012).

However, most previous studies have not considered the different materiality of ESG issues, which might explain the mixed results of prior research. Khan et al. (2016) first introduced the concept of materiality in the context of CSP research. According to their definition, information about ESG issues is likely to be material if it potentially affects an investor's investment (or voting) decision. I follow Khan et al. (2016) by considering the distinction between material and non-material CSP as highly relevant in the realm of CSP. Previous research on

[16] Please refer to the sections I.2 and I.1 for further reference as to how the terms CSR, CSI, CSP, ESG performance, and ESG risks are used in this thesis. Also note that the first paper of this thesis—per default—only refers to *positive* CSP (measured by ESG performance) as opposed to *negative* CSP.

EQ has neglected this distinction. This study attempts to fill this gap by mainly examining whether distinguishing between material and non-material ESG issues makes a difference to EQ. Therefore, I recalculate Thomson Reuters ESG scores based on the materiality guidance of SASB.

I use a sample of 542 European firms from 27 industries and implement two common proxies for earnings quality: AQ and REM. For the AQ proxy, first findings show that only material CSP firms exhibit significantly higher EQ. For the REM proxy, however, results indicate that both material and non-material CSP firms manipulate real operating activities more strongly and thus exhibit lower EQ. I control for regulated and non-regulated industries in robustness checks and find that the difference between material and non-material CSP firms does not exist in regulated industries.

This study contributes to the literature in five ways. First, although some research on the relation between CSP and EQ exists, no single study has considered the concept of materiality. I use materiality guidance from SASB to add explanatory power to the model investigating the relationship between CSP and EQ. Second, by recalculating Thomson Reuters ESG scores based on the guidance from SASB, I provide insights into the operationalization and reliability of the materiality guidance from SASB for academic research. Third and to the best of my knowledge, no previous study on earnings quality has considered extensive information about ESG issues from Thomson Reuters. Most research has relied on data from KLD, which, according to previous literature (e.g., Humphrey et al., 2012), have some significant drawbacks related to their binary nature. Thomson Reuters ESG data are in a percentile-ranking format. Fourth, I investigate a European sample of 542 firms from 27 different industries throughout 12 years. Grewatsch and Kleindienst (2017) and McWilliams et al. (2006) pointed out that going beyond the US market is essential since European companies very likely differ from US companies in terms of investments in ESG issues. Fifth, in robustness checks, this study considers industry regulation as a potential additional factor.

The remainder of the study is structured as follows. Section II.2 discusses the previous literature and develops the hypothesis. Section II.3 states the methodology and data of this study. Section II.4 presents the results. Section II.5 concludes.

II.2 Theoretical Background, Literature Review, and Hypothesis Development

This section discusses the importance of materiality in the domain of CSP (II.2.1). I further briefly define the key theories and terms of this study and discuss the literature on CSP and EQ (II.2.2). Finally, I derive the hypotheses of this study (II.2.3).

II.2.1 Corporate Social Performance and Materiality

Social and environmental issues like climate or demographical change are becoming increasingly important in the corporate world. Hence, firms disclose more and more information about their performance on ESG issues. For instance, for reasons of reputation, many firms seek to perform well across as many ESG issues as possible. As a result, the amount of reported information about ESG issues per firm is continuously increasing. Even though, Grüner et al. (2014) link reporting about ESG issues to lower stock return volatility and lower negative abnormal returns, they also find that reporting about ESG issues has increased substantially between 2002 and 2012. In the EU, the non-financial reporting directive[17] mandatorily requires larger companies to disclose information about their ESG issues. Hinze and Freidank (2018) posit that this will even further increase reporting and based on the current reporting system, the standalone CSR report is probably the best alternative to comply. Most CSR reports, however, have long since passed the one hundred-page mark.

One of the concerns accompanying this development is that researchers and practitioners have treated firms from different industries almost as one homogeneous unit. However, firms from different industries operate under distinct environmental, social, governmental, and financial conditions (Baird et al., 2012; Grewatsch and Kleindienst, 2017; Schreck, 2011). ESG issues are very likely to be unequally crucial for firms in different industries. Some ESG issues might be beneficial to a firm and its shareholders, whereas others could be detrimental or have no effect at all. For instance, previous studies do not agree on whether corporate governance has a positive affect on CSR reporting quality (Dienes and Velte, 2014). Attig et al. (2013) provide

[17] Directive 2014/95/EU. Available via: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32014L0095.

suggestive evidence that managers should only promote investments in and report about those ESG issues that are directly linked to their primary stakeholders. In other words, the materiality of ESG issues for investors very likely varies from industry to industry.[18]

I account for this different materiality of ESG issues in this study. Khan et al. (2016) introduced a materiality classification based on guidance from SASB. They argued that investments in material sustainability issues and not investments in non-material[19] sustainability issues contribute to firm performance. In fact, Khan et al. (2016) state that investments in non-material sustainability issues have either no value or even negative value (i.e., reducing firm value). They find that the effect of CSP on firm performance depends on whether a firm performs well on material or non-material ESG issues. The finding of Khan et al. (2016) suggests that future research should consider materiality when investigating CSP. Similarly, findings of other studies suggest that considering materiality might be essential due to different effects of the components of CSP on firm performance (e.g., Aerts et al., 2008; Derwall et al., 2011; Ibikunle and Steffen, 2017). I use materiality guidance from SASB to classify ESG issues into material or non-material for each industry. For the materiality classification process, see section II.3.2. To get an overview of SASB's activities, see Appendix A.1.

II.2.2 Empirical Literature on Earnings Quality

Following Dechow et al. (2010), I define earnings quality as follows: "Higher quality earnings provide more information about the features of a firm's financial performance that are relevant to a specific decision made by a specific decision-maker." This definition implies that investors and analysts can make better investment decisions and recommendations based on earnings of higher quality, rendering EQ as an indispensable key figure. Furthermore, the earnings management behavior of managers negatively influences EQ. Thus, if managers manipulate the earnings

[18]For instance, a healthcare firm is more likely to be affected by product safety issues than a utility firm, which is more likely to be concerned by environmental issues, like CO2 emissions.

[19]Thanks to an external piece of advice, I use the term "non-material"—in contrast to Khan et al. (2016) who used the term "immaterial". I justify this adjustment by the slightly different philosophical meaning between information that is just "non-material" and information that is explicitly "immaterial." More specifically, "immaterial" is not the exact opposite of "material," while "non-material" is. SASB only explicitly determines which ESG issues are "material" and not the ones which are "immaterial." Therefore, the residual ESG issues not considered in their standard setting process are rather "non-material" than explicitly "immaterial."

of a firm more strongly, EQ suffers. Moreover, I define the term corporate social performance as a firm's performance on ESG issues in a given year.

A couple of studies has empirically examined the relationship between CSP and EQ. Results of these studies are mixed. Some research has posited that firms with high CSP exhibit higher EQ. Most of these studies are based on ethical theories (e.g., Carroll, 1979; Garriga and Melé, 2004; Mackey et al., 2007), reputational theories (e.g., Fombrun and Shanley, 1990) and financial theories (e.g., Edmans, 2011; Feng et al., 2015; Ferrell et al., 2016; Khan et al., 2016; Kiessling et al., 2015) of CSP. In general, such theories indicate a positive relation between CSP and EQ. What follows outlines empirical studies that support these theories by finding that firms with high CSP also exhibit higher EQ.

Kim et al. (2012) and Hong and Andersen (2011) used data from KLD and a US sample. Overall, their results indicated that all the proxies of EQ they used (e.g., accruals quality and real earnings management) positively relate to CSP. Moreover, the finding of Kim et al. (2012) that high CSP firms also experience less accounting and auditing enforcement releases supports a positive relationship between CSP and EQ.[20] Hong and Andersen (2011) used alternative measures of CSP, but this did not change the outcome.

Labelle et al. (2010) focused on diversity management, which is a specific dimension of CSP. They used a rather small sample of 78 firms and hypothesized that diversity management reduces opportunities for earnings management. Labelle et al. (2010) found that firms with stronger diversity management indeed show higher EQ. Their evidence coincides with Hong and Andersen (2011) and Kim et al. (2012), at least as the diversity dimension of CSP is concerned.

Gao and Zhang (2015) found that firms with high CSP demonstrate a stronger relationship between actual returns and reported earnings. They also showed that actual returns correlate more strongly with future returns for high CSP firms compared to low CSP firms. Gao and Zhang (2015) even stated that CSP "adds a unique 'quality dimension' to earnings attributes."

[20] Since 1999, the US Securities and Exchange Commission publishes online accounting and auditing enforcement releases. These releases relate to financial reporting enforcement actions against managing boards of firms. I will not use them as a proxy for earnings management since accounting and auditing enforcement releases are US-specific, while this study uses a European sample.

Other studies, however, have suggested that firms with high CSP exhibit less earnings quality. Most of these studies are based on self-interest or instrumental theories of CSP. They suggest that a manager's self-interest might incentivize to use CSP as an instrument to reach personal goals. According to Hegarty and Sims (1978), the unethical behavior of managers always accumulates whenever sufficient incentives exist to act in one's interest. Incentives to act according to self-interest might also lead managers to manipulate earnings more strongly to reach their personal goals. At the same time, these managers might use CSP as a disguise from their destructive earnings manipulation behavior. What follows outlines studies that support such self-interest and instrumental theories of CSP and indicate that high CSP firms exhibit lower EQ.

The first study, Chih et al. (2008), investigated an international sample of 1,653 firms. They used three proxies for earnings quality: earnings smoothing, loss avoidance, and earnings aggressiveness. In contrast to the findings of Hong and Andersen (2011) and Kim et al. (2012), Chih et al. (2008) found that CSP increases earnings aggressiveness. In other words, CSP seems to lower EQ—at least if EQ is proxied by earnings aggressiveness. However, CSP is positively associated with EQ if earnings smoothing or loss avoidance are used as proxies. Chih et al. (2008) stated that the mixed results might be due to different regulatory and legal enforcement environments across the countries of the sample firms. Therefore, I included country fixed effects (FE) in my model to control for any differences between countries.[21] Moreover, as put forward in Dechow et al. (2010), all three proxies used in Chih et al. (2008) might be no suitable measures for earnings management.

Kim and Venkatachalam (2011) examined whether a significantly higher information risk (i.e., low EQ) justifies higher financial returns of "sin stocks."[22] They based their study on previous findings of superior financial returns of "sin stocks" Hong and Kacperczyk (2009). Kim and Venkatachalam (2011) hypothesized that "sin stocks" only offer superior returns because they also have poorer financial reporting quality (i.e., low EQ). Surprisingly, they found

[21] In this case, using a FE design is appropriate since I explicitly do not intend to explain differences between countries in this study.

[22] The term "sin stocks" usually refers to firms that belong to the tobacco, gaming, alcohol, arms & weapons, and munition industries. The exact definition depends on factors like geography, culture, regulatory environment, etc. "Sin stocks" are usually also referred to as non-ethical and as low performers on different kinds of ESG issues.

superiority in the financial reporting of "sin stocks". This finding suggests that firms belonging to a sin industry exhibit higher EQ than non-"sin stocks". However, pertaining to a "sin industry" is not a discretionary factor (Kim et al., 2012). Therefore, the evidence of Kim and Venkatachalam (2011) does not rule out that high CSP firms could not also perform better on EQ.

Prior et al. (2008) posited that firms might use CSP to obscure earnings management. Their sample consisted of 593 firms, which they examined between 2002 and 2004. They found a positive relationship between CSP and earnings management. However, their finding was only significant for regulated firms. Unregulated firms did not exhibit more EQ.

Similarly to Prior et al. (2008), Martínez-Ferrero et al. (2016) empirically investigated whether CSP is strategically used to mask earnings management. Their hypothesis was twofold: First, CSP might lead to a decrease in cost of capital and an increase in reputation; and second, the reduction in cost of capital and the increase in reputation cancel out the negative effects of earnings management. They showed that firms could offset negative effects of earnings management on the cost of capital and reputation by committing to CSP. This result suggests that managers might intentionally use CSP to obscure the negative effects of earnings management, resulting in lower EQ. Drawing on Martínez-Ferrero et al. (2016), I consider a corresponding hypothesis for a negative impact of CSP on EQ later on.

This study builds on earlier studies and yet is clearly distinct many respects. For example, both Hong and Andersen (2011) and Kim et al. (2012) used data from KLD and a US sample to measure CSP. In contrast, this study used data on European firms which very likely differ with regard to investments into CSP (Grewatsch and Kleindienst, 2017; McWilliams et al., 2006). Moreover, Chih et al. (2008) and Jordaan et al. (2018) used index membership as the sole proxy for CSP.[23] Using index membership is prone to criticism because it does not allow any variability between the firms' degree of CSP. In contrast, a percentile-ranking system—as used in this study—accounts for different levels of CSP between firms. Furthermore, Jordaan et al. (2018) focused on a small South African sample of 214 observations.

[23] If a firm was a member of the FTSE4Good Global Index—a well-known sustainability index—Chih et al. (2008) classified this firm as a socially responsible one. If a firm was not a member of the FTSE4Good Global Index, they classified it as a non-socially responsible firm.

Most importantly, however, no previous study has considered the different materiality of ESG issues. Even though, recent research on CSP has shown that distinguishing between material and non-material ESG issues potentially matters (e.g., Khan et al., 2016). Therefore, this study builds on this evidence by taking into account the materiality of ESG issues in the relationship between CSP and EQ. The following section develops the hypotheses to test the relation between material and non-material CSP and EQ.

II.2.3 Hypotheses Development

On the first level, this study tries to shed more light on the general relationship between CSP and EQ by using a European sample of 5,974 firm-year observations over the years 2005 to 2016. In line with previous studies (e.g., Hong and Andersen, 2011; Kim et al., 2012), my first hypothesis is based on ethical theories (Carroll, 1979; Garriga and Melé, 2004) and reputational theories (e.g., Fombrun and Shanley, 1990) of CSP. Ethical theories implicate that firms have a certain societal and environmental responsibility towards other stakeholders apart from shareholders and beyond just making profits. Firms would have to acquire a certain legitimacy towards their stakeholders. Managers of firms with good CSP potentially acknowledge this notion and thus are more likely to restrain themselves from managing earnings. They are therefore more committed to a transparent financial reporting than managers of firms that do not attach such great importance to CSP. Similar to ethical theories, reputational theories point to a firm's incentive to not only comply with the rules. Beyond that, firms have to behave legitimately to not damage their reputation and therefore their future success. Based on these two theory streams and consistent with Kim et al. (2012), I state the *Transparency Hypothesis*:[24]

$H1_A$**:** Firms with high CSP exhibit more earnings quality than firms with low CSP

On the contrary, self-interest and instrumental theories of CSP (e.g., Fritzsche, 1991; Hegarty and Sims, 1978; Hemingway and Maclagan, 2004) indicate that firms with good CSP could exhibit more earnings management (i.e., less EQ). Managers that are driven by self-interest and moral hazard might abuse the benefits of CSP to distract from other harmful activities.

[24] I control for financial motives of CSP by including a financial performance measure in the research design (Edmans, 2011; Khan et al., 2016; Waddock and Graves, 1997).

Managers may obscure the negative effects of manipulating earnings by showing a good performance on corporate social and environmental issues. This notion comes very close to "greenwashing" according to which CSP would be just a disguise. Consequently, CSP might negatively correlate with EQ. Such a negative hypothesis has gained empirical support by Prior et al. (2008) and more recently by Martínez-Ferrero et al. (2016). Consistent with Kim et al. (2012), I state the *Instrumental Hypothesis*:[25]

$H1_B$**:** Firms with high CSP exhibit less earnings quality than firms with low CSP[26]

On the second level, I extend the relationship between CSP and EQ by the materiality dimension. To the best of my knowledge, not a single study on EQ has considered materiality in a European context. However, distinguishing the importance of ESG issues between industries is an emerging topic in CSP research (cf., section II.2.1). Khan et al. (2016) showed that it is crucial to distinguish between material and non-material ESG issues in the corporate financial performance dimension. The distinction between material and non-material ESG issues may also explain some of the inconsistent findings of previous research on EQ.

Managers of firms that perform well on material ESG issues (henceforth, material CSP firms) may take corporate social and environmental activities and reporting more seriously. Like the *Transparency Hypothesis* ($H1_A$), this notion is based on ethical and reputational theories of CSP. Firstly, managers of material CSP firms care more about determining which ESG issues are material and which are non-material to their firm and investors. Secondly, managers of material CSP firms might care more about performing well across exactly these material ESG issues. The intrinsic motivation of managers to perform well especially on those ESG issues that are material to investors may also affect their behavior in the manipulation of earnings. Such managers are likely to be more transparent in financial reporting, to have more discipline in managerial decision-making, and to present their firm's financial results more accurately. In line with Kim et al. (2012), Hong and Andersen (2011) and Khan et al. (2016), I predict that the positive relationship between CSP and EQ is even stronger for material CSP firms and state

[25] For a more detailed discussion on the theories related to the *Transparency Hypothesis* and the *Instrumental Hypothesis*, see Kim et al. (2012).

[26] For both hypotheses $H1_A$ and $H1_B$, the corresponding null hypothesis, $H1_0$, is that there is no significant relation between CSP and EQ at all.

the *Materiality Hypothesis*:

$H2_a$**:** Material CSP firms exhibit more earnings quality than non-material CSP firms

Managers of firms that primarily perform well on non-material ESG issues (henceforth, non-material CSP firms), however, may follow a different strategy. Based on self-interest and instrumental theories of CSP (e.g., Fritzsche, 1991; Hegarty and Sims, 1978; Hemingway and Maclagan, 2004), non-material CSP firms might try to distract from earnings management by performing well on as many ESG issues as possible. Be it material or non-material ESG issues. The vast majority of ESG issues is non-material across all industries, as my classification of Thomson Reuters ESG issues has shown (cf., section II.3.2). In other words, in each industry, only a minority of ESG issues is material. Since non-material ESG issues are more numerous, managers are more likely to distract from earnings management by performing well on these non-material ESG issues. If a non-material CSP firm uses performance on non-material ESG issues to distract from earnings management, EQ should be lower compared to firms that do not have this strategy. In line with Khan et al. (2016), Martínez-Ferrero et al. (2016), and Prior et al. (2008), I therefore state the *Immateriality Hypothesis*:[27]

$H2_b$**:** Non-material CSP firms exhibit less earnings quality than material CSP firms[28]

II.3 Methodology & Data

This section is crucial to preparing the empirical analysis and consists of five sub-sections. Sub-section II.3.1 characterizes the ESG data and sample of this study. Sub-section II.3.2 exemplifies the classification of ESG data into material and non-material. Sub-section II.3.3 demonstrates how the ESG scores are recalculated based on the previous materiality classification. Sub-section II.3.4 states the proxies for EQ. Sub-section II.3.5 develops the empirical model and specifies the control variables.

[27] In robustness checks, I test whether the results for H2 from the main analysis (i.e., full sample) also hold after controlling for the regulation of industries (i.e., sub-sample).

[28] For both hypotheses $H2_a$ and $H2_b$, the corresponding null hypothesis, $H2_0$, is that there is no significant relation between material or non-material CSP and EQ at all.

II.3.1 ESG Data

I used Thomson Reuters ESG scores to measure the CSP of a European sample.[29] I focused on the European market as previous literature has mostly investigated US samples. Thus, the findings of previous studies are not applicable to European firms. This study closes this gap by examining a purely European sample. In 2017, Thomson Reuters ESG covered over 6,500 firms globally. However, if only considering Europe, this number decreased to about 1,400 firms. Previous research has widely used the predecessor of Thomson Reuters ESG scores, ASSET4 ratings (e.g., Eccles et al., 2014; Feng et al., 2015; Ferrell et al., 2016; Liang and Renneboog, 2017).

Thomson Reuters ESG scores measure a firm's CSP relative to industry peers across ten ESG categories: Resource Use, Emissions, Innovation, Workforce, Human Rights, Community, Product Responsibility, Management, Shareholders, and CSR Strategy. These categories include 406 different ESG issues in total, for which data goes back to 2002. Scoring involves a percentile-ranking system. Theoretically, a score above .99 indicates that a firm belongs to the top 1% performing firms in that specific category or issue. In contrast, a score below .01 indicates that a firm belongs to the bottom 1% performing firms.

Compared to Thomson Reuters ESG data, previous research has relied mostly on KLD data. However, previous studies have mentioned several concerns that arise from using KLD data (e.g., Humphrey et al., 2012). Thomson Reuters ESG data appear to mitigate some of these concerns. Especially two characteristics of Thomson Reuters ESG data are worth mentioning: First, in contrast to KLD data, Thomson Reuters ESG scores are not binary; and second, Thomson Reuters ESG scores account for the positive and negative management of ESG issues of firms. Thus, Thomson Reuters ESG data allow greater variability in ESG issues between firms by having a percentile-ranking system instead of binary scores. Also, Thomson Reuters ESG scores adequately reflect good (or bad) management of ESG issues of firms operating in environmentally- or socially-sensitive industries. By default, KLD data punishes such firms

[29] Thomson Reuters ESG scores are an enhanced version of ASSET4 ratings. The new Thomson Reuters scores have replaced ASSET4 ratings in 2017. Thomson Reuters has retired ASSET4 ratings in 2018. Detailed information about the methodology of Thomson Reuters and the enhancements compared to ASSET4 scores is available via: https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/esg-scores-methodology.pdf

despite all their potential and well-intentioned efforts to manage ESG issues.

The sampling process of this study has been subject to three restricting factors related to Thomson Reuters ESG data: First, I focused on Europe; thus, only firms headquartered on the European continent were included in the sample. Second, only firms which had ESG scores available for at least six consecutive years were included. This restriction was necessary to ensure an adequate time-series. Together, these two screening criteria reduced the sample size to 872 firms. Finally, financial sector firms were excluded given that their earnings differ fundamentally from earnings of non-financial firms (Hong and Kacperczyk, 2009; Kim et al., 2012). This exclusion was essential to provide adequate comparability of results and reduced the sample to 719 firms.

Moreover, Thomson Reuters ESG scores had to be recalculated to integrate the materiality guidance from SASB into ESG scores. Recalculation followed the same calculation procedure as that of Thomson Reuters, except for the inclusion of the materiality dimension. This calculation procedure uses industry benchmarks to calculate ESG category scores that are comparable within industries. The size of the industry benchmark (=peer group) depends on the industry. For some industries, data were available for 30 firms, while for other industries data were only available for five firms. The minimum number of firms with available data was set to ten firms per industry. Otherwise, this study did not consider the industry in its sample. This exclusion was necessary to ensure an adequate benchmarking and recalculation process of ESG scores. However, this strict approach further reduced the sample by 151 firms, resulting in a sample of 565 non-financial firms from 27 different industries. The investigation period was 2005 to 2016. Excluding 256 firm-year observations with missing control variable data finally resulted in 5,974 firm-year observations (542 firms).[30] Table A.1 in the appendix summarizes the firm-year observations and the number of industries in the sample.

[30]The exclusion of 256 firm-year observations due to missing control variable data did not influence the benchmarking process of Thomson Reuters ESG scores. Please refer to the next two sub-sections for details on the recalculation and benchmarking process of ESG scores.

II.3.2 Materiality Classification Approach of ESG Issues

Thomson Reuters carefully selects a subset of 178 ESG issues from an entire 406 ESG issues for its ESG scoring process. Thomson Reuters bases its selection on criteria such as comparability, data availability, and industry relevance.[31] This study, however, used the guidance of SASB to determine which ESG issues are likely to be material or non-material to investors for each industry.[32] Compared to the Thomson Reuters industry relevance criteria, the guidance of SASB goes much further in assessing which ESG issues are relevant and which are not. Given that every ESG issue could be potentially material to some industry, I refrained from excluding specific ESG issues ex-ante. Excluding ESG issues before applying the materiality guidance would have biased the recalculation procedure.

For instance, SASB classifies the ESG topic "Water Management" as material for the industry "Oil & Gas – Exploration & Production." Therefore, the ESG issues "Water Use to Revenues USD," "Water Withdrawal Total," "Fresh Water Withdrawal Total," and "Water Recycled" (from the "Resource Use" category) were classified as material for that industry. Another example is the "Pharmaceutical" industry, for which SASB classifies "Corruption & Bribery" as a material ESG topic. Accordingly, ESG issues related to corruption or bribery were classified as material. All other topics that were not classified as material are non-material for a specific industry. In general, some industries only have a few material ESG issues overall, whereas others have numerous material ESG issues concentrated in certain ESG categories (e.g., "Resource Use").

Surprisingly, merging the ESG data and the materiality guidance showed that all of the 123 issues from the three governance categories (i.e., Management, Shareholders, and CSR Strategy) are non-material across all industries. One possible reason for this might be that SASB labels governance issues differently than Thomson Reuters (e.g., "Leadership and Governance" instead of "Management"). Further, Thomson Reuters covers specific corporate governance-related issues in the environmental or social dimension, whereas SASB covers them in separate

[31] For more comprehensive information, please visit: https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/esg-scores-methodology.pdff

[32] Henceforth, material and non-material mean "likely to be material" and "likely to be non-material," respectively.

dimensions related to corporate governance.

Khan et al. (2016) found that the subjectivity of researchers only slightly biased the materiality classification process. In fact, discrepancies between researchers were found just in 2% of cases in their study. Hence, it is reasonable to assume that if subjectivity in the classification of ESG issues exists, its extent is likely to be negligible. Eventually, material ESG issues were used to recalculate material CSP, while non-material ESG issues were used to recalculate non-material CSP, as shown in the next section.

II.3.3 Recalculation of the Materiality-Based ESG Scores

This study used an adjusted Thomson Reuters methodology to recalculate materiality-based ESG scores. Hence, I followed Thomson Reuters by calculating industry-benchmarked percentile scores for each ESG issue and category according to the following percentile formula:[33]

$$\frac{\#\text{ companies with worse value} + \frac{\#\text{ companies with the same value incl. the current one}}{2}}{\#\text{ companies with a value}} \tag{II.3.1}$$

The data for ESG issues are either numerical or Boolean. If a numerical ESG issue had no value (=0), it was not considered in calculating the percentile ESG issue and category scores. According to Thomson Reuters, it would be highly inaccurate to consider a numerical ESG issue with no data available (=0) in the calculation. Boolean ESG issues take either the value one or the default of zero.

The ESG category scores are the equally weighted sum of all relevant ESG issues in the respective industry. The percentile ESG issue and category scores are insensitive to outliers because they are based on the rank of firms within the same industry. The percentile ESG category scores were calculated by using the same percentile formula from above (II.3.1). The number of ESG issues determined the weight of each category. Categories with more ESG

[33] The scoring procedure of Thomson Reuters is described in detail in: https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/esg-scores-methodology.pdf

issues had higher weights than those with fewer ESG issues.

The ESG dimension scores (Environmental, Social, and Governance) were calculated as the weighted average of their category scores. The Environmental dimension consists of three categories: Resource Use, Emissions, and Innovation. The Governance dimension also has three categories: Management, Shareholders, and CSR Strategy. The Social dimension consists of four categories: Workforce, Human Rights, Community, and Product Responsibility.

Finally, overall non-material and material ESG scores were calculated as the weighted average of the three dimension scores. Once again, the number of relevant (i.e., non-material and material) ESG issues determined the weights of each dimension. However, as mentioned before, no single governance issue has been classified as material according to the SASB guidance. Therefore, the material ESG score only includes the scores from the Environmental and Social dimensions.

The two recalculated non-material and material ESG scores were used to measure non-material CSP and material CSP. They were calculated on firm level and for each year between 2005 and 2016. Figure A.1 in the appendix illustrates the procedure.

II.3.4 Proxies for Earnings Quality

This section discusses the selection of different proxies for earnings quality. Dechow et al. (2010) have reviewed several proxies used in previous research. I used the following two proxies: accruals quality (i.e., abnormal accruals) and the manipulation of real earnings (i.e., REM). These two proxies are most suited to measure earnings management and have thus been used by previous research (Hong and Kacperczyk, 2009; Kim et al., 2012; Prior et al., 2008).

The first model, a modified version of the accruals quality model (Francis et al., 2005), measures earnings management through abnormal accruals. Using the Dechow and Dichev (2002) approach would not be sufficient. As Francis et al. (2005) stated, the Dechow and Dichev (2002) approach does not account for the sources of measurement error in accruals quality. Nor does it answer the question of whether accruals quality is higher or lower due to intentional

earnings management, management failure, or the management environment (Francis et al., 2005). As this study's objective is mainly concerned with the discretionary part of accruals, I used the modified accruals quality model.

The traditional accruals quality model accounts for the fact that the Dechow and Dichev (2002) approach only measures current accruals (Francis et al., 2005). The traditional accruals quality model measures abnormal accruals by including cash flow variables and fundamental information from financial reporting. As shown in McNichols (2002), including both cash flow variables and fundamental balance sheet variables increases explanatory power. The modified accruals quality model of this study further integrates prior-year ROA to control for performance (Kim et al., 2012; Kothari et al., 2005). AQ is measured in absolute terms, as in Kim et al. (2012), to account for the possibility of negative and positive accruals. Higher values of AQ indicate more earnings management and thus lower EQ. Lower values of AQ indicate less earnings management and, hence, higher EQ (cf., Figure A.2 in Appendix A.6). The modified accruals quality model is derived in Appendix A.4.

The second model uses measures for REM as an appropriate proxy for EQ. More specifically, I used the model of Roychowdhury (2006), which adopts three measures for REM. Academic research widely acknowledges the validity of this model (Hong and Andersen, 2011; Jordaan et al., 2018; Kim et al., 2012). It uses abnormal operating cash flow (ACFO), abnormal production cost (APROD), and abnormal discretionary expense (ADIEXP) as proxies for REM. The three measures were calculated by regressing sales from financial reporting on actual operating cash flow, production costs, and discretionary expense. The residuals (ε) of the three regressions measure the level of REM. High residual values of ACFO and ADIEXP indicate less REM and thus higher EQ (cf., Figure A.3 in Appendix A.7). In contrast, high residual values of APROD indicate more REM and thus lower EQ (cf., Figure A.4 in Appendix A.7).

Further, a combined measure of REM (COMBREM) was included (Cohen et al., 2008; Cohen and Zarowin, 2010; Kim et al., 2012). The combined measure aggregated the three proxies above (ACFO, APROD, and ADIEXP) by calculating ACFO minus APROD plus ADIEXP. High residual values of COMBREM indicate less real earnings management and thus higher EQ (cf., Figure A.3 in Appendix A.7). The derivation of all four models (ACFO, APROD, ADIEXP,

and COMBREM) to measure real earnings management can be found in Appendix A.5.

II.3.5 Empirical Model and Control Variables

I estimated the following cross-sectional time-series regression model for different specifications to test the hypotheses mentioned above:[34]

$$\begin{aligned} EarningsQuality_{j,t} = {} & \alpha_0 + \beta_1 CSPScore_{j,t} + \beta_2 EQSubstitute_{j,t} \\ & + \beta_3 WorkingCapital_{j,t} + \beta_4 RDActivities_{j,t} + \beta_5 CGScore_{j,t} \\ & + \beta_6 Age_{j,t} + \beta_7 MKTB_{j,t-1} + \beta_8 PriceEarnings_{j,t-1} \\ & + \beta_9 ROA_{j,t-1} + \beta_{10} Leverage_{j,t-1} + \beta_{11} EOIndicator_{j,t} + \varepsilon_{j,t} \end{aligned} \tag{II.3.2}$$

where:

$EarningsQuality_{j,t}$, depending on the specification, is either firm *j*'s measure of accruals quality (AQ, PositiveAQ, or NegativeAQ) or firm *j*'s measure of real earnings management (ACFO, APROD, ADIEXP, or COMBREM) in year *t*

$CSPScore_{j,t}$, depending on the specification, is either firm *j*'s overall CSP, firm *j*'s recalculated material CSP score, or firm *j*'s recalculated non-material CSP score in year *t*

$EQSubstitute_{j,t}$ is either firm *j*'s accruals quality (AQ) calculated by the modified accruals quality model based on Francis et al. (2005) and Kothari et al. (2005), or firm *j*'s REM measured by the combined variable (COMBREM) in year *t*

$WorkingCapital_{j,t}$ is firm *j*'s working capital in year *t* scaled by lagged total assets

$RDActivities_{j,t}$ is firm *j*'s expenses for R&D scaled by sales in year *t*

[34] The standard model was run on a within-firm level with industry dummies to control for FE. However, running the model on a within-industry level did not substantially change the results.

$CGScore_{j,t}$ is firm j's Thomson Reuters governance score in year t

$Age_{j,t}$ is the number of years of firm j since its IPO plus two years

$MKTB_{j,t-1}$ is firm j's market-to-book ratio lagged by one year

$PriceEarnings_{j,t-1}$ is firm j's price-earnings ratio lagged by one year

$ROA_{j,t-1}$ is firm j's industry-adjusted return on assets (profitability) lagged by one year

$Leverage_{j,t-1}$ is firm j's long-term debt (excl. accounts payable) lagged by one year and scaled by lagged total assets

$EOIndicator_{j,t}$ is an indicator variable that takes the value of 1 if firm j has a next year $(t+1)$ equity offering which is publicly known in year t

In accordance with Zang (2012) and Kim et al. (2012), regressions of model II.3.2 comprise a substitute variable to control for a substitutive relation of different earnings management measures. In regressions using AQ as the dependent variable, COMBREM is included as a control variable. Vice versa, in regressions using REM measures (like COMBREM) as the dependent variable, AQ is included as a control variable.

Several control variables are implemented to prevent concerns about correlated omitted variables. Working capital controls for liquidity because firms that face liquidity risk are more likely to engage in earnings management (Park and Shin, 2004). Working capital is defined as current assets minus current liabilities. R&D intensity potentially affects CSP and earnings (McWilliams and Siegel, 2000). Therefore, R&D expense is included to control for a firm's R&D activity (Kim et al., 2012). Following Kim et al. (2012), it has been controlled for the effect of corporate governance on financial reporting. The corporate governance score of Thomson Reuters proxies for the different corporate governance intensities of firms.

Moreover, a variable for firm age controls for a potential effect of a firm's changing maturity on financial reporting (Kim et al., 2012). Roychowdhury (2006) indicated that growth opportunities might affect earnings management. Thus, I include measures for growth opportunities (market-to-book ratio (MKTB) and price-earnings ratio). I further add ROA to the model

to isolate the effect of material CSP on EQ from the impact of financial performance on EQ. Following Kim et al. (2012), Kim and Park (2005), and Teoh et al. (1998), a measure for leverage and an indicator variable for equity offerings are included. They are included to control for effects of debt- and next year equity-offering-related incentives on EQ. Where applicable, variables are scaled by total assets or sales. All continuous control variables are winsorized at the 1st and 99th percentile to control for the outlier effect.

II.4 Empirical Results

This section is split into three parts. Sub-section II.4.1 shows the descriptive statistics of the full sample. Sub-sections II.4.2 and II.4.3 present the results from the cross-sectional time-series regressions. Among other things, Sub-section II.4.4 discusses additional analyses of sub-samples to check for the robustness of results.

II.4.1 Descriptive Statistics

Table 4 provides descriptive statistics for the full sample. More specifically, it reports cross-sectional time-series mean and standard deviation (SD), the 25th percentiles, the median, and the 75th percentiles. The means, the 25th percentiles, the medians, and the 75th percentiles of material CSP and non-material CSP are very close to each other. The standard deviations of both variables are also relatively close to each other. These similar characteristics suggest that there is not much variability between the two variables. The lack of variability indicates that it might be difficult to detect differences between the effect of material CSP and the effect of non-material CSP on EQ. Moreover, the correlation between overall CSP and non-material CSP is even more strikingly high (.98). However, this is not very surprising, as most of the Thomson Reuters ESG issues were non-material according to my materiality assessment and thus included in both the overall CSP and non-material CSP measures. Therefore, it is only consequential that the non-material CSP score correlates strongly with the overall CSP score.

Descriptive statistics further show that the sample firms are 21 years of age on average, which is comparable to the 22 years in Kim et al. (2012). Moreover, sample firms have an

average P/E ratio of 19.1 and a ROA of 6.7%. Both mean values seem reasonable for the cross-section of industries. The MKTB of sample firms is slightly higher as in Kim et al. (2012) (2.9 compared to 2.4) but still in a reasonable range.

Table 5 shows pairwise correlations among selected dependent, independent, and control variables. Overall CSP significantly ($p < 0.01$) and negatively correlates with AQ, indicating that CSP firms exhibit higher accruals quality (i.e., earnings quality). Material CSP and non-material CSP negatively correlate with AQ, but the coefficient is only significant ($p < 0.01$) for non-material CSP. The correlation coefficients have the same sign and the correlation between material CSP and non-material CSP is relatively high (.56) and significant ($p < 0.01$). Both findings support the lack of variability between material CSP and non-material CSP. This lack of variability is the first indication that the Materiality Hypothesis ($H2_a$) may not hold. Moreover, CGScore is significantly and positively correlated with overall CSP, material CSP, and non-material CSP. This positive correlation indicates that firms that perform better on environmental and social issues also show stronger performance on corporate governance issues.

Surprisingly, industry-adjusted ROA positively correlates with material CSP (significant at $p < 0.01$) but not with non-material CSP. This bivariate correlation suggests that firms with a higher material CSP also have a higher return on assets. However, bivariate correlations should be taken with caution since strong correlations between the independent and control variables exist. Therefore, the multiple regression model in the next sub-section accounts for a host of firm characteristics and includes standard errors that were adjusted by a two-dimensional cluster, as well as country and industry FE.

Table 4: Summary Statistics

Dependent Variables	N	Mean	SD	25th	Median	75th
AQ	5974	0.204	0.211	0.051	0.144	0.283
PositiveAQ	3462	0.259	0.217	0.095	0.201	0.362
NegativeAQ	1886	−0.155	0.120	−0.226	−0.125	−0.058
ACFO	5974	0.046	0.271	−0.051	0.019	0.144
APROD	5974	−0.053	0.371	−0.202	−0.026	0.141
ADIEXP	5974	−0.032	0.163	−0.054	0.005	0.030
COMBREM	5974	0.065	0.387	−0.147	0.028	0.235
Independent Variables						
Overall CSP	5974	0.497	0.167	0.378	0.511	0.625
Material CSP	5974	0.494	0.190	0.356	0.502	0.634
Non-Material CSP	5974	0.498	0.175	0.367	0.509	0.634
Control Variables						
WorkingCapital	5974	0.104	0.182	0.000	0.079	0.190
RDActivities	5974	0.021	0.044	0.000	0.001	0.020
CGScore	5974	0.502	0.204	0.353	0.505	0.654
Age	5974	21.476	26.091	5.000	14.000	25.000
MKTB	5974	2.898	3.104	1.123	2.036	3.645
PriceEarnings	5974	19.129	25.051	8.168	14.696	21.489
ROA	5974	0.067	0.074	0.017	0.054	0.104
Leverage	5974	0.188	0.144	0.074	0.171	0.277
EOIndicator	5974	0.188	0.391	0.000	0.000	0.000

Notes: This table reports time-series averages of cross-sectional mean and SD, the 25th percentiles, the 50th percentiles (median), and the 75th percentiles for all dependent, independent and control variables. AQ represents the absolute value of abnormal accruals, measured by the residuals of the modified accruals quality model (AQM) (Francis et al., 2005). PositiveAQ comprises only firm-year observations with positive accruals. NegativeAQ comprises just firm-year observations with negative accruals. PositiveAQ and NegativeAQ sum up to the full sample of 5,794 observations. ACFO is the level of abnormal operating cash flow. APROD is the level of abnormal production cost, where production cost is measured as the sum of cost of goods sold (COGS) and change in inventory. ADIEXP is the level of abnormal discretionary expense, where discretionary expense is measured as the sum of R&D expense, advertising expense, and selling, general & administrative expense (SG&A) expense. COMBREM is the sum of real earnings management proxies, measured as ACFO – APROD + ADIEXP. Overall CSP is the overall firm-specific performance on all environmental and social issues, measured by TR ESG score. Material CSP is the firm-specific performance on material environmental and social issues, measured by the recalculated TR ESG score, only considering the material ESG issues per industry. Non-Material CSP is the firm-specific performance on non-material environmental and social issues, measured by the recalculated TR ESG score, just considering the non-material ESG issues per industry. WorkingCapital is current assets net of current liabilities scaled by lagged total assets. RDActivities is expenses for R&D scaled by sales. CGScore is performance on corporate governance issues, measured by TR governance score (the higher the score, the stronger the performance). Age is the number of years of a firm since its IPO plus two years. MKTB is the price per share divided by book value per share. PE is price per share divided by earnings per share. ROA is mean industry-adjusted return on assets, measured by income before extraordinary items divided by total assets. Leverage is long-term debt (excl. accounts payable) scaled by total assets. EOIndicator is an indicator variable whether a firm has a next year equity offering (i.e., either takes the value of 1 or 0).

Table 5: Pairwise Correlations

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	AQ	1																
2	ACFO	0.24*	1															
3	APROD	-0.15*	0.25*	1														
4	ADIEXP	-0.13*	-0.12*	0.13*	1													
5	COMBREM	0.26*	0.34*	-0.71*	0.22*	1												
6	Overall CSP	-0.04*	-0.01*	0.03*	-0.04*	-0.06*	1											
7	Material CSP	-0.02	-0.01	0.05*	-0.05*	-0.08*	0.68*	1										
8	Non-Material CSP	-0.04*	-0.01	0.03*	-0.03*	-0.05*	0.98*	0.56*	1									
9	WorkingCapital	0.15*	0.12*	-0.02	-0.03*	0.08*	-0.08*	-0.05*	-0.09*	1								
10	RDActivities	0.16*	0.20*	-0.03*	-0.25*	0.07*	0.05*	0.02	0.05*	0.17*	1							
11	CGScore	0.05*	-0.01	-0.02	-0.04*	0.00	0.23*	0.18*	0.22*	-0.06*	0.04*	1						
12	Age	-0.01	0.04*	0.07*	0.05*	-0.02	0.16*	0.11*	0.15*	0.05*	0.02	0.12*	1					
13	MKTB	0.15*	0.15*	-0.03*	-0.10*	0.08*	0.01	0.03*	0.00	0.05*	0.15*	0.01	0.03*	1				
14	PriceEarnings	0.02	0.03*	0.02	0.01	0.01	0.01	0.03*	0.01	0.01	0.05*	-0.01	0.00	0.11*	1			
15	ROA	0.32*	0.27*	-0.12*	-0.13*	0.24*	0.02	0.04*	0.01	0.23*	0.15*	0.05*	0.02	0.38*	0.00	1		
16	Leverage	-0.12*	-0.10*	-0.06*	0.02	0.00	0.06*	0.03*	0.06*	-0.30*	-0.18*	0.02	-0.05*	-0.11*	-0.03	-0.27*	1	
17	EOIndicator	-0.09*	-0.06*	-0.02	0.01	-0.02	0.06*	0.03*	0.06*	-0.13*	-0.03	0.03*	-0.02	-0.12*	-0.02	-0.10*	0.15*	1

Notes: This table reports univariate pairwise Pearson correlations among the dependent, independent, and control variables. For variable descriptions, see Table 4. * indicates statistical significance at least at the 5% level.

II.4.2 The Relation between Material and Non-Material CSP and Accruals Quality

Table 6 presents the results of the multiple regression model for accruals quality. The cross-sectional model included country and industry FE to control for differences between countries and industries. The standard errors used to calculate the test statistics and significance levels were adjusted by a two-dimensional cluster to account for possible correlation of residuals across firm and time (Cameron et al., 2011; Gow et al., 2010; Thompson, 2011). Accruals are measured in absolute terms and lower values of AQ indicate higher EQ (c.f., section II.3.4).

As reported in Table 6, the coefficients of overall CSP and non-material CSP are negative but insignificant ($p > 0.05$). This finding is not very surprising as both variables have large overlaps—as shown in the descriptive statistics. Hence, there is no support for either the *Transparency Hypothesis* ($H1_A$) or *Instrumental Hypothesis* ($H1_B$), which state that CSP should be positively or negatively related to EQ, from this regression.

On the other hand, firms performing well on material ESG issues, display statistically significant higher accruals quality (i.e., earnings quality). The coefficient of material CSP is also larger than that of overall CSP or non-material CSP. In fact, a one standard deviation increase in material CSP leads to a decrease of 3.4% in the manipulation of accruals, all else constant. These findings lend support to the Materiality Hypothesis $H2_a$. Managers of material CSP firms are, indeed, less likely to engage in earnings management, according to the accruals quality measure.

The fact that there is not much variability between material and non-material CSP firms speaks against this result and therefore makes the finding of a positive relation even more robust. This result is also in line with Hong and Andersen (2011) and Kim et al. (2012). However, as I find for a European sample, the positive relation does only hold for material CSP firms. After including country FE, the results of Chih et al. (2008) for a negative relationship between CSP and EQ could not be supported. Moreover, in contrast to Prior et al. (2008), the findings of this study so far do not indicate that CSP negatively affects EQ.

I arrived at these results after controlling for REM by the combined measure of REM proxies, COMBREM. The coefficients of COMBREM for the AQ regressions are positive and

Table 6: Regressing Accruals Quality on Material and Non-Material CSP

	Accruals Quality		
Independent Variables	(1)	(2)	(3)
Overall CSP	−0.028		
	(−1.76)		
Material CSP		−0.038**	
		(−3.01)	
Non-Material CSP			−0.029
			(−1.93)
Control Variables			
COMBREM	0.090**	0.089**	0.090**
	(9.32)	(9.18)	(9.34)
WorkingCapital	0.114**	0.114**	0.114**
	(5.35)	(5.38)	(5.33)
RDActivities	0.746**	0.746**	0.747**
	(7.35)	(7.37)	(7.36)
CGScore	0.010	0.011	0.010
	(0.81)	(0.90)	(0.82)
Age	0.000**	0.000**	0.000**
	(−4.28)	(−4.31)	(−4.27)
PriceEarnings	0.000	0.000	0.000
	(−0.07)	(−0.00)	(−0.08)
MKTB	0.003**	0.003**	0.003**
	(4.04)	(4.08)	(4.03)
ROA	0.318**	0.320**	0.318**
	(6.61)	(6.67)	(6.62)
Leverage	0.038*	0.038*	0.038*
	(2.10)	(2.10)	(2.11)
EOIndicator	−0.014*	−0.014*	−0.014*
	(−2.56)	(−2.56)	(−2.56)
Constant	−0.032	−0.033	−0.031
	(−0.74)	(−0.77)	(−0.72)
Country FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
N	5974	5974	5974
Adjusted R2	0.18	0.18	0.18

Notes: This table shows coefficients of the multiple regressions of absolute accruals quality (AQ) on Overall CSP, Material CSP and Non-Material CSP. All specifications include country and industry fixed-effects. For variable descriptions, see Table 4. t statistics in parentheses. * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

significant ($p < 0.01$). Higher values of COMBREM indicate less real earnings management and thus higher earnings quality (cf., Figure A.3 in Appendix A.7), whereas higher values of AQ indicate more earnings management and thus less earnings quality (cf., Figure A.2 in Appendix A.6). Therefore, positive coefficients suggest that firms are using real earnings management as a substitute for earnings management through accruals, and the other way around. This finding is in line with the results of Kim et al. (2012), Zang (2012), and Cohen et al. (2008) who also suggest a substitutive relation between real earnings management and earnings management through accruals.

Turning to control variables, I found that firms that are more profitable show higher EQ (i.e., significant and negative association between industry-adjusted ROA and AQ). Furthermore, firms that spend more on R&D also exhibit significantly ($p < 0.01$) higher accruals (i.e., less earnings quality). This finding corresponds to the notion that firms use expenses on research and development to manipulate earnings. This notion is also supported by the REM regressions as shown in the next section.

II.4.3 The Relation between Material and Non-Material CSP and Real Earnings Management

Table 7 reports the results of the multiple regression model for REM on overall CSP. Again, I include country and industry FE and adjust the standard errors by a two-dimensional cluster to calculate the test statistics and significance levels.

Turning to the overall CSP regressions, the coefficients are significant ($p < 0.01$ and $p < 0.05$) for all REM proxies (ACFO, ADIEXP, APROD, and COMBREM). However, the direction of the correlation between overall CSP and REM is in contrast to Hong and Andersen (2011) and Kim et al. (2012). Kim et al. (2012) found for all REM proxies that CSP is associated with higher EQ. In contrast, I found for all REM proxies that CSP is associated with lower EQ. The coefficients for abnormal operating cash flow, ACFO, abnormal discretionary expense, ADIEXP, and the combined measure of real earnings management, COMBREM, are negative. Lower values of ACFO, ADIEXP, and COMBREM indicate more real earnings management and thus lower earnings quality (cf., Figure A.3 in Appendix A.7). For abnormal production

costs, APROD, the coefficient is positive. Higher values of APROD indicate more real earnings management and thus i.e., lower earnings quality (cf., Figure A.4 in Appendix A.7).

In fact, a one standard deviation increase in overall CSP leads to an increase of 7.4% in the manipulation of real operating activities, all else constant. Finding a positive relation between CSP and REM lends support to the *Instrumental Hypothesis* ($H1_B$). Managers of firms performing well on ESG issues tend to distract from harmful activities such as earnings management by engaging in CSP. This result might be partially driven by the different geographical focus of this study compared to previous ones (Europe vs. United States). It also corresponds to the findings of Martínez-Ferrero et al. (2016) and Prior et al. (2008) who found evidence that managers disguise earnings management by engaging in CSR. It seems that high CSP firms do not manipulate accruals more strongly than low CSP firms (cf., previous section). However, they seem to engage in earnings management by manipulating real operating activities more strongly.

Turning to materiality, Table 8 shows that the Materiality Hypothesis ($H2_a$) does not hold for proxies of REM. Coefficients of both material CSP and non-material CSP show negative signs for the three REM proxies ACFO, ADIEXP, and COMBREM, and positive signs for the REM proxy APROD (significant at $p < 0.01$ and $p < 0.05$). The coefficients' signs are no different from the overall CSP regressions. In other words, no difference exists between CSP firms, material CSP firms, and non-material CSP firms in their behavior when manipulating real operating activities. As indicated by descriptive statistics, this finding might hold due to the lack of variability in the ESG data of Thomson Reuters.

Table 7: Regressing Real Earnings Management on Overall CSP

	ACFO	APROD	ADIEXP	COMBREM
Independent Variable	(1)	(2)	(3)	(4)
Overall CSP	−0.075**	0.069**	−0.016*	−0.172**
	(−5.34)	(3.32)	(−2.23)	(−6.69)
Control Variables				
AQ	0.124**	−0.152**	−0.010	0.274**
	(6.25)	(−6.87)	(−1.32)	(9.49)
WorkingCapital	0.037*	−0.079**	0.002	0.099**
	(2.11)	(−2.86)	(0.21)	(3.12)
RDActivities	0.293**	−0.559**	0.770**	1.785**
	(2.61)	(−4.76)	(7.96)	(9.53)
CGScore	0.005	−0.037*	−0.030**	0.008
	(0.45)	(−2.17)	(−4.36)	(0.37)
Age	0.000**	0.000	0.000	0.000
	(5.50)	(1.32)	(1.58)	(1.62)
PriceEarnings	0.000	0.000	0.000	0.000
	(0.24)	(−0.15)	(1.54)	(0.95)
MKTB	0.003**	−0.003*	0.000	0.007**
	(4.16)	(−2.03)	(1.05)	(4.45)
ROA	0.447**	−0.658**	−0.078**	1.034**
	(10.15)	(−11.51)	(−3.21)	(14.09)
Leverage	0.030	−0.249**	−0.045**	0.233**
	(1.80)	(−9.27)	(−4.99)	(7.48)
EOIndicator	−0.008	0.013	−0.005	−0.029**
	(−1.44)	(1.63)	(−1.73)	(−3.01)
Constant	−0.503**	0.262**	−0.094**	−0.831**
	(−7.07)	(4.12)	(−10.87)	(−7.61)
Country FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
N	5974	5974	5974	5974
Adjusted R2	0.58	0.43	0.27	0.21

Notes: This table shows coefficients of the multiple regressions of proxies for real earnings management (REM) on Overall CSP. All specifications include country and industry fixed-effects. For variable descriptions, see Table 4. t statistics in parentheses. * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

Table 8: Regressing Real Earnings Management on Material and Non-Material CSP

	ACFO		APROD		ADIEXP		COMBREM	
Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Material CSP	−0.047**		0.090**		−0.016**		−0.161**	
	(−4.08)		(5.22)		(−2.86)		(−7.79)	
Non-Material CSP		−0.071**		0.052**		−0.015*		−0.147**
		(−5.44)		(2.58)		(−2.24)		(−5.99)
Control Variables								
AQ	0.124**	0.124**	−0.150**	−0.153**	−0.010	−0.010	0.271**	0.275**
	(6.25)	(6.26)	(−6.74)	(−6.89)	(−1.37)	(−1.32)	(9.34)	(9.52)
WorkingCapital	0.041*	0.037*	−0.080**	−0.080**	0.003	0.002	0.106**	0.100**
	(2.32)	(2.09)	(−2.92)	(−2.89)	(0.27)	(0.21)	(3.31)	(3.14)
RDActivities	0.272*	0.295**	−0.555**	−0.554**	0.768**	0.770**	1.756**	1.781**
	(2.43)	(2.62)	(−4.74)	(−4.72)	(7.94)	(7.97)	(9.37)	(9.51)
CGScore	−0.001	0.004	−0.039*	−0.034*	−0.030**	−0.030**	0.003	0.004
	(−0.08)	(0.40)	(−2.33)	(−2.00)	(−4.47)	(−4.39)	(0.13)	(0.18)
Age	0.000**	0.000**	0.000	0.000	0.000	0.000	0.000	0.000
	(5.04)	(5.50)	(1.32)	(1.43)	(1.47)	(1.57)	(1.35)	(1.49)
PriceEarnings	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.35)	(0.22)	(−0.26)	(−0.14)	(1.61)	(1.53)	(1.13)	(0.93)
MKTB	0.003**	0.003**	−0.003*	−0.003*	0.000	0.000	0.007**	0.007**
	(4.26)	(4.16)	(−2.12)	(−2.04)	(1.09)	(1.04)	(4.62)	(4.46)
ROA	0.445**	0.447**	−0.662**	−0.657**	−0.078**	−0.078**	1.036**	1.033**
	(10.11)	(10.15)	(−11.58)	(−11.50)	(−3.18)	(−3.21)	(14.11)	(14.09)
Leverage	0.029	0.031	−0.248**	−0.249**	−0.045**	−0.044**	0.232**	0.234**
	(1.75)	(1.82)	(−9.23)	(−9.28)	(−4.98)	(−4.99)	(7.41)	(7.50)
EOIndicator	−0.008	−0.008	0.013	0.013	−0.005	−0.005	−0.029**	−0.029**

Table 8 continued from previous page

	(−1.47)	(−1.43)	(1.61)	(1.63)	(−1.73)	(−1.73)	(−3.02)	(−3.01)
Constant	−0.507**	−0.501**	0.264**	0.263**	−0.095**	−0.094**	−0.838**	−0.829**
	(−7.08)	(−7.07)	(4.12)	(4.12)	(−10.82)	(−10.85)	(−7.60)	(−7.61)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	5974	5974	5974	5974	5974	5974	5974	5974
Adjusted R2	0.58	0.58	0.43	0.43	0.27	0.27	0.21	0.21

Notes: This table shows coefficients of the multiple regressions of proxies for real earnings management (REM) on Material CSP and Non-Material CSP. All specifications include country and industry fixed-effects. For variable descriptions, see Table 4. t statistics in parentheses. * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

The substitutive relation between AQ and REM also supports this conclusion. As mentioned, I included AQ as a substitutive measure for REM in this section's regressions. The coefficients of AQ of this section's regressions confirm the substitutive relationship. The coefficient signs of AQ suggest that AQ and REM act as substitutes in all of the eight regressions of REM proxies in Table 8. These results again indicate that sample firms might use the manipulation of real operating activities instead of accruals to manage earnings.

Turning to control variables, industry-adjusted ROA is significantly ($p < 0.01$) and positively (negatively) correlated with ACFO and COMBREM (APROD) (Table 8). This finding suggests that firms that are more profitable also show higher EQ (cf., Figure A.3 in Appendix A.7). However, if I use ADIEXP, more profitable firms show less EQ (negative coefficient, significant at $p < 0.01$; Table 8).

II.4.4 Additional Analyses and Robustness Checks

I used an alternative proxy for accruals quality as a robustness check. Hence, operating cash flow (CFO) was removed from the modified accruals quality model as in Kim et al. (2012) and Kothari et al. (2005). Consequently, total current accruals (TCA) only mapped into change in revenue (REV), property, plant, and equipment (PPE), and ROA. Unreported results show that this alternative proxy did not change the results substantially.

Furthermore, I performed two sub-sample tests. The first sub-sample test only comprised firm-years with positive and negative numbers for accruals quality (PositiveAQ, NegativeAQ). As expected, the signs of the coefficients from the PositiveAQ regressions were negative, whereas the signs of the coefficients from the NegativeAQ regressions turned positive. The direction of signs is in line with the results from my main analysis. However, three out of four coefficients were statistically insignificant. The results from the additional sub-sample tests can be found in Table A.2 in Appendix A.8.

In the second sub-sample test, I separated firms in regulated industries (i.e., electric utilities, oil & gas, and telecommunications) from firms in non-regulated industries by following Prior et al. (2008). Antitrust authorities are exerting greater pressure and surveillance on man-

agers of firms in regulated industries to comply with current regulations concerning price fixing and market shares (Prior et al., 2008). On the one hand, firms in regulated industries have less accounting discretion and are therefore less likely to manage earnings (Kim et al., 2012). The pressure from antitrust authorities, on the other hand, might trigger earnings management because managers want their firm look less profitable (Watts & Zimmerman, 1978). Consistent with Kim et al. (2012), using the sub-sample of regulated firms, the results from the absolute AQ regression show that only firms in regulated industries exhibit a statistically significant higher EQ. Moreover, the relationship is even stronger than in the full sample.

As Table 9 shows, now the coefficients of both material CSP and non-material CSP are negatively and significantly ($p < 0.01$) correlated with absolute AQ (Columns 1 and 2). As lower values of AQ indicate less earnings management (i.e., higher earnings quality), this finding is in line with the result for the full sample. In fact, coefficients are substantially higher than in the full-sample regressions. Consequently, the positive relation between CSP and EQ seems to be even stronger for regulated firms. However, the difference in earnings management through accruals between material CSP and non-material CSP firms does not hold for regulated firms. Moreover, my finding is in contrast to Prior et al. (2008) who found that regulated firms are more likely to manage earnings than non-regulated ones. The fact that more regulated firms have lower accounting discretion (Kim et al., 2012) likely explains my result of a higher earnings quality.

When investigating only non-regulated firms, however, I find that CSP is not significantly correlated with absolute accruals quality anymore (Table 9; Columns 3 and 4). As in Prior et al. (2008), the results are only significant for regulated firms. This finding suggests that regulated firms most likely drive the result of a positive relation between CSP and EQ in the full sample regression.

Moreover, the results for the substitutive relation between REM and AQ holds for both sub-samples (regulated and non-regulated firms). COMBREM is positive and significant ($p < 0.01$) in all four regression for both sub-samples (Table 9). This finding suggests that both regulated and non-regulated sample firms are using real earnings management and the manipulation of accruals in a substitutive manner.

Table 9: Analysis of Sub-Samples of Firms from Regulated and Non-Regulated Industries

	Accruals Quality			
	Regulated Firms		Non-Regulated Firms	
Independent Variables	(1)	(2)	(3)	(4)
Material CSP	−0.133**		−0.017	
	(−3.63)		(−1.33)	
Non-Material CSP		−0.244**		0.025
		(−6.76)		(1.56)
Control Variables				
COMBREM	0.124**	0.135**	0.082**	0.083**
	(3.27)	(3.69)	(8.06)	(8.20)
WorkingCapital	−0.032	−0.035	0.130**	0.133**
	(−0.35)	(−0.37)	(6.11)	(6.25)
RDActivities	1.000	1.268	0.799**	0.781**
	(1.33)	(1.72)	(7.82)	(7.61)
CGScore	−0.058	−0.025	0.019	0.012
	(−1.90)	(−0.82)	(1.50)	(0.91)
Age	−0.002**	−0.002**	0.000**	0.000**
	(−5.63)	(−5.42)	(−2.64)	(−2.96)
PriceEarnings	0.000	0.000	0.000	0.000
	(0.19)	(0.30)	(−0.36)	(−0.37)
MKTB	0.005*	0.004	0.002**	0.002**
	(2.01)	(1.50)	(2.81)	(2.73)
ROA	0.408**	0.429**	0.294**	0.288**
	(3.31)	(3.56)	(5.68)	(5.57)
Leverage	0.132*	0.099	0.003	0.001
	(2.46)	(1.91)	(0.14)	(0.07)
EOIndicator	−0.044**	−0.039**	−0.007	−0.007
	(−3.67)	(−3.40)	(−1.11)	(−1.10)
Constant	0.389**	0.436**	−0.067	−0.069
	(8.75)	(9.97)	(−1.53)	(−1.57)
Country FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
N	1078	1078	4896	4896
Adjusted R^2	0.31	0.33	0.17	0.17

Notes: This table shows coefficients of the multiple regressions of absolute accruals quality (AQ) on Material CSP and Non-Material CSP for the sub-samples of regulated and non-regulated firms. All specifications include country and industry fixed-effects. For variable descriptions, see Table 4. t statistics in parentheses. * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

Even though I used logged and scaled variables, tests for non-stationarity were performed for the cross-sectional time series data to check for unit roots. Fisher type tests (Choi, 2001) showed that all variables are stationary (i.e., do not have a unit root), except for AGE.[35] In other words, due to stationarity, the likelihood of having spurious regression results is low. The results of the alternative proxy and the Fisher type tests are unreported but available on request.

II.5 Concluding Remarks

This study has investigated whether firms with different performance on material and non-material ESG issues also show disparities in how they engage in earnings management. On the one hand, I found no support for the *Transparency Hypothesis* ($H1_A$). In general, CSP is not statistically and positively related to EQ. On the other hand, first evidence indicated that material CSP firms exhibit higher EQ according to the accruals quality proxy. This finding is in support of the *Materiality Hypothesis* ($H2_a$). In robustness checks, however, I showed that the difference in EQ between material and non-material CSP firms does not exist when controlling for the regulation of industries. In regulated industries, both material and non-material CSP firms exhibit higher EQ (supporting the *Transparency Hypothesis* $H1_A$ for this type of firms). In non-regulated industries, however, no such significant relationship exists at all. In line with Prior et al. (2008), the positive relation between CSP and EQ is probably solely driven by regulated firms in the full sample. Although I found that, according to the AQ proxy, regulated firms do not show more earnings management (as in Prior et al. (2008)) but less.

I further investigated the manipulation of real operating activities and found that both material and non-material CSP firms exhibit more REM (i.e., lower earnings quality). This finding lends support to the *Instrumental Hypothesis* ($H1_B$), indicating that firms might use CSP to distract from their manipulation of real operating activities. Results of this study are different from Kim et al. (2012) and Hong and Andersen (2011) who found that CSP is negatively associated with REM. However, it is in line with Martínez-Ferrero et al. (2016) and Prior

[35] By its nature, the variable Age has a non-constant mean. However, two-way clustering of standard errors across firm and year potentially mitigates concerns of spurious regression results due to non-stationarity.

et al. (2008), finding that firms use CSP to disguise from harmful activities such as earnings management. Moreover, this study did not find any evidence in support of the *Immateriality Hypothesis* ($H2_b$).

Results for both REM and AQ seem to contradict each other only at first glance. I further found in robustness checks that both regulated and non-regulated firms might use real earnings management as a substitute for manipulating accruals. Results from previous research support this finding (Cohen et al., 2008; Kim et al., 2012; Zang, 2012). This study's results taken together strongly suggest that, on the one hand, CSP firms in regulated industries exhibit higher EQ due to less manipulation of accruals. On the other hand, however, instead of manipulating accruals, CSP firms seem to engage in the manipulation of real operating activities.

Further, this study's overall results do not support the notion that material CSP firms do perform any better or worse on earnings quality than non-material CSP firms. The baseline result of a positive relation between material CSP and EQ could not be substantiated when controlling for regulated and non-regulated industries. However, it seems that industry membership (regulated vs. non-regulated) should not be neglected in CSP research as its effects on the results are potentially substantial.

This study is subject to at least three limitations. First, one drawback is applying guidance from the SASB on a European sample. SASB received more feedback in its standards-setting work from US organizations than European organizations. Although the SASB standards are in use by investors and companies globally, their accounting guidance is more often focused on US regulations than European regulations. While this focus is likely to be unproblematic concerning the investors—the preferences of global investors do not differ substantially between Europe and the United States; it might pose certain inconsistencies concerning disclosure activities of firms. That means different legal and cultural aspects influence a firm's disclosure activity, and corporate disclosure differs between the United States and Europe. I did not account for a legal or cultural moderating factor on a firm's disclosure activity.

Second, even though I controlled for a host of firm characteristics that potentially affect earnings quality, I was not able to cover all aspects. Some of these aspects would be difficult to measure. For instance, companies might use their discretion regarding carbon dioxide (CO_2)

transfer prices to manipulate earnings.

Third, many measures of CSP (i.e., ESG scores) lack standardization and harmonization in general, which makes it difficult to draw conclusions about other measures of CSP to a certain degree. Connected to this point, the lack of variability between material and non-material CSP is possibly the reason why differences between the two are difficult to detect. Even though Thomson Reuters has implemented a percentile scoring system for their ESG scores, which I adopted for this study, differences between the overall material and non-material CSP scores appear to be rather small.

Future research should use materiality guidance that is better adapted to the European market and European standards. Standard setting organizations other than SASB (e.g., Global Reporting Initiative or the German DNK) or privately organized business intelligence providers (e.g., eRevalue) have emphasized the importance of materiality. The availability of more comprehensive and more precise materiality guidance is just a matter of time. Future research should benefit by using guidance on materiality that better fits European characteristics and idiosyncrasies. Future studies could also focus more on the differences between regulated and non-regulated industries, as regulation appears to play a role. Another avenue for future research should be investigating how the variability in ESG data can be increased. More variability in ESG data would likely reveal new insights not only with regard to earnings quality but also in CSP research in general.

III Paper 2: Risk Mitigation of Corporate Social Performance in US Class Action Lawsuits[36]

Daniel V. Fauser, Sebastian Utz

Abstract We investigate the relationship between corporate social performance and litigation risk by examining US class action lawsuits. Previous research indicates that CSP has risk management characteristics and produces insurance-like effects in corporate crises. Therefore, we posit that CSP serves as a risk management tool against class action lawsuits. We find that a one standard deviation improvement in ESG controversies of an average sample firm reduces litigation risk from 3.1% to 2.4%. Moreover, an average sample firm with low ESG performance exhibits losses twice as high in market value compared to a firm with high ESG performance, i.e., an abnormal loss of US $1.14bn. Implementing our findings with a long-only trading strategy yielded monthly alphas up to 1.27% in the period from 2014 to 2017, suggesting that investors benefit from lower litigation risk and insurance-like protection. Our findings imply that market participants benefit from lower litigation risk and insurance-like protection when investing in socially and environmentally responsible firms. An extensive series of robustness checks further corroborates our results. Our results have important implications for the market participants' risk assessment of investment targets.

Key Words: Litigation risk; corporate social performance; investment strategy; risk management; ESG; event study

[36] We are grateful to Prof. Dr. Bert Scholtens, Prof. Dr. Timo Busch, Prof. Dr. Judith Walls, and the seminar participants at Maastricht University, University of Liechtenstein, the Corporate Finance Day in Groningen, and the Annual Conference of the Global Research Alliance for Sustainable Finance and Investment (GRASFI) in Oxford, for their very valuable feedback and comments. The paper has been accepted for publication in the Financial Analysts Journal. The journal is ranked as *B* in VHB-JOURQUAL3, and its impact factor was 1.816 in 2018. We thank the journal's executive editor, Prof. Stephen J. Brown, Ph.D., and Co-Editor, Daniel Giamouridis, Ph.D., as well as two anonymous reviewers for their suggestions, which have significantly improved our paper. Slight adjustments to the paper have been made for the purpose of this thesis (e.g., additional tables, hypotheses).

III.1 Introduction

"Dieselgate" and Volkswagen's systematic and blatant fraud shook the public in 2015. The emissions scandal was described by numerous renowned media as the most outrageous corporate fraud of the 21st century. On the one hand, such corporate scandals may seriously impact customers (e.g., overpaying for cars) and the public (e.g., externalized environmental costs). On the other, corporate scandals and managerial misconduct entail considerable financial (e.g., compensation payments) and reputational risks. Managerial misconduct, which often culminates in serious corporate scandals, might be driven by low standards of corporate environmental sustainability and social responsibility. Research on corporate social responsibility has therefore repeatedly studied the link between a firm's environmental, social, and corporate governance engagement—so-called CSP—and firm value and risk (e.g., Aouadi and Marsat, 2018; Bouslah et al., 2013).

CSP is a multidimensional construct that combines diverse firm behavior regarding environmental (e.g., pollution control), social (e.g., diversity), and corporate governance (e.g., stakeholder strategy) issues (Waddock and Graves, 1997).[37] Research indicates that CSP has risk-reducing characteristics (Cai et al., 2016; Harjoto and Laksmana, 2018). From an investment perspective, a major motivation to consider information on CSP is the relevance to investment performance and firm risk (Amel-Zadeh and Serafeim, 2018; Bouslah et al., 2013). One important aspect of firm risk is determined by the probability of extreme events with substantial negative consequences. We investigate US class action lawsuits as examples of such extreme events. Previous research has neglected the potential explanatory power of ESG controversies and performance in the context of class action lawsuits. ESG controversies represent negative CSP, while ESG performance represents positive CSP.

Therefore, we investigate (1) to what extent past ESG controversies and ESG performance predict future class action lawsuits and (2) whether ESG performance mitigates the negative financial impact of class action lawsuit filings on firm value. Further, we synthesize the findings on these objectives in a trading strategy whether predicted litigation risk and ESG per-

[37] In our theoretical part, the term CSP refers to a firm's positive and negative activities in all dimensions of sustainability and social responsibility (i.e., environmental, social, and governance). In our models, we use two measures for positive and negative CSP: (1) ESG performance to proxy for positive CSP; (2) ESG controversies to proxy for negative CSP.

formance have asset pricing implications.

Our study reinforces the notion of positive and negative CSP as indicators of both transparency and risk management policies. Measures for CSP offer insights into firms that cannot be derived from financial reporting and thus increase transparency. The disclosure of CSP measures may reduce stakeholders' search and evaluation costs for such non-financial information (Kennett, 1980). Moreover, engaging in CSP has been perceived as a risk management initiative (Cai et al., 2016; Harjoto and Laksmana, 2018). Bouslah et al. (2013) have supported this notion of risk management theory. In line with that notion, we use ESG controversies as a proxy for *weak* risk management policies and ESG performance as a proxy for *strong* risk management policies.

Furthermore, class action lawsuits have negative effects on short- and long-term performance (Bauer and Braun, 2010). We posit that once a class action lawsuit has occurred, positive CSP mitigates the negative effects on firm value because it acts as a "reservoir of goodwill." By generating this reservoir, firms can build up moral capital among different stakeholders and thus insure themselves against negative events (Fombrun and Shanley, 1990) and against negative impacts on performance during a corporate crisis (Godfrey, 2005; Koh et al., 2014). Thus, firms that consider the interests of all stakeholders achieve long-term financial success as these interests are aligned (Freeman, 2010). For instance, the financial performance of firms benefited from high positive CSP during the 2007–08 financial crisis (Lins et al., 2017) and in cases of corrupt firm behavior (Hong et al., 2019). Recently, however, Liu et al. (2020) found contrary evidence: Positive CSP does not protect shareholder wealth in environmental lawsuits. Moreover, only weak evidence exists that CSP systematically reduces credit risk (Stellner et al., 2015). The previous discussion suggests that research is still far from conclusive as to whether positive CSP can protect against losses from legal disputes.

Anecdotal evidence for this potential insurance-like protection is displayed by two healthcare firms (Mallinckrodt and TherapeuticsMD) from our sample. Both faced a class action lawsuit due to neglecting their reporting duties regarding a monopolistic position and a new drug application, respectively.[38] Both lawsuits were filed in early 2017 but impacted firm value

[38] More information on examples of class action lawsuits that are included in our sample can be found in Table B.2.

substantially differently. While TherapeuticsMD lost about 40% of its firm value in 21 days around the lawsuit filing, Mallinckrodt lost only about 9%. One potential reason for this discrepancy was certainly the severity of the claim, which was possibly linked to ESG standards in the respective firm. At this point, CSP comes into play. According to an external evaluation of its CSP and compared to its peers, Mallinckrodt was among the top-performing firms, while TherapeuticsMD was at the bottom of the CSP spectrum.

We add empirical evidence for CSP's risk management potential based on a panel data set containing 7671 firm-year observations of US firms from 2003 to 2017. We use scores from Thomson Reuters to proxy for firm exposure towards ESG controversies and ESG performance. Our measure for litigation risk is based on class action lawsuits from the Securities Class Action Clearinghouse (SCAC) of Stanford Law School. We explain litigation risk in an extended multiple panel regression model based on Kim and Skinner (2012) as a function of ESG controversies, ESG performance, and well-documented control variables. We find that a one standard deviation improvement in our measure for ESG controversies of an average firm reduces the probability for a class action lawsuit filing by 22.6%, i.e., from 3.1% to 2.4%.

We further apply an event study to investigate the insurance-like effects of ESG performance during litigation. Hawn et al. (2018) considered event study methodology as a powerful tool for investigating the link between CSP and stock market reactions. We measure the market reaction to a firm's distress originating from litigation by abnormal returns (ARs). ARs are calculated in two ways. First, based on actual returns from CRSP and on the expected returns from Fama and French's (1993) three-factor model. Second, by applying a propensity score matching approach using actual returns of the treated firms (i.e., firms facing a class action lawsuit) and of the matched control firms. We find evidence of an insurance-like effect of positive CSP for firms facing class action lawsuits. In particular, an average sample firm (market cap: US $22.1bn) with low ESG performance experiences an abnormal loss in market value of US $1.14bn around the filing of the lawsuit compared to a high ESG performance firm.

We contribute to the discussion on the validity of CSP regarding risk management in a litigation risk context. First, our study quantifies how a firm's ESG controversies directly inform about the risk of litigation. We show that running a business responsibly can reduce the number

of litigation cases and thus prevent reputational damage and higher cost of capital. Accordingly, socially and environmentally responsible businesses are also of interest to investors as the tail event risk of litigation is lower. Second, we add to research on CSP, firm value, and risk (e.g., Harjoto and Laksmana, 2018; Bouslah et al., 2013). In particular, we provide evidence for a moderating effect of positive CSP on firm value during class action lawsuits. Third, our findings are especially relevant to investors due to the potential asset pricing implications. We implement a long-only trading strategy that focuses on those firms with the lowest litigation risk and the highest ESG performance. Following this trading strategy does not only decrease an investor's exposure to idiosyncratic risk but also provides an outperformance compared to asset pricing models. More specifically, the trading strategy yields positive alphas up to 0.73% (before transaction costs). The respective breakeven transaction costs of 0.45% are clearly above the transaction costs that institutional investors usually pay. Thus, our alphas remain positive even after considering transaction costs, indicating that our proposed trading strategy is feasible.

Our contribution to related studies on the relationship between CSP, risk, and firm performance (e.g., Hong et al., 2019; Koh et al., 2014; Aouadi and Marsat, 2018) is fourfold: First and foremost, none of these studies has investigated the prediction of litigation risk as a function of positive and negative CSP. Second, we used actual litigation events (i.e., class action lawsuits, though not particularly ESG-related) instead of using firms from low and high litigation industries as proxies (e.g., Koh et al., 2014). Our study also differs from Hong et al. (2019) by focusing on a more comprehensive indicator of corporate misconduct compared to only corruption-related corporate misconduct. Whereas Aouadi and Marsat (2018) focused on ESG controversies and firm value, we investigate the insurance-like protection of firm value through positive CSP in a litigation context. Third, we derive the actual market impact on firm value around class action lawsuit filings. In contrast, Koh et al. (2014) adopted financial and accounting performance metrics in a residual income model. Moreover, we use a homogeneous US sample of class action lawsuits compared to a cross-country sample (e.g., Aouadi and Marsat, 2018). Fourth, none of the above studies has developed a trading strategy that attempts to implement their obtained findings in a profitable way.

III.2 Corporate Social Performance and the Probability of Litigation

Supported by the growing evidence on CSP's investment relevance, investors are increasingly considering CSP when assessing a firm's financial performance and risk. While numerous empirical studies examined the direct relationship between CSP and financial performance (e.g., Ferrell et al., 2016; Aouadi and Marsat, 2018), others focused on the risk dimension and the potential risk-reducing role of CSP (e.g., Harjoto and Laksmana, 2018; Bouslah et al., 2013).

The risk-reducing effect of CSP is based on two arguments from risk management theory: First, increased disclosure about CSP decreases information asymmetries (Cui et al., 2018). Reduced information asymmetries due to ESG engagement translate into lower risk (Cai et al., 2016). Second, engaging in CSP might itself be perceived as a risk management initiative by providing necessary tools for dealing with risks. For example, CSP is considered to mitigate the risk-taking behavior of managers by working as a control mechanism (Harjoto and Laksmana, 2018). In fact, the majority of evidence suggests that CSP decreases idiosyncratic risk and total firm risk (e.g., see Bouslah et al., 2013; Utz, 2018).

However, research on the relationship between CSP and litigation risk is rather scarce, despite its potential implications for investment portfolios. To the best of our knowledge, no previous study has investigated whether involvement in ESG controversies informs about a firm's probability of facing litigation. However, ESG controversies may inform outsiders about misconduct concerning a company's products, customer health & safety, and shareholder rights, amongst others. Such misconduct may be a sign of bad risk-taking behavior and weak ESG risk management (see above), and thus is potentially an early indicator of higher litigation risk. Hence, and in accordance with Chen (2016), we argue that the occurrence of ESG controversies indicates negative CSP and therefore higher litigation risk. For instance, so-called "sin firms" have been shown to suffer disproportionately under litigation (Hong and Kacperczyk, 2009). From an investor's perspective, ESG controversies might help to explain litigation risk—a part of idiosyncratic portfolio risk.

Based on CSP's risk management characteristics (Cai et al., 2016; Harjoto and Laksmana, 2018; Benlemlih et al., 2018), we posit that higher ESG performance, indicating stronger

positive CSP, is linked to lower litigation risk. We further posit that fewer ESG controversies, indicating weaker negative CSP, are linked to lower litigation risk (Chen, 2016). Therefore, we formally state the two hypotheses:[39]

$H1_A$: Higher positive CSP (i.e., a stronger ESG performance) is linked to a lower current-year probability of facing a class action lawsuit

$H1_B$: Lower negative CSP (i.e., fewer previous-year ESG controversies) is linked to a lower current-year probability of facing a class action lawsuit

The second part of our paper investigates the potential moderating effects of positive CSP on firm value in cases where a class action lawsuit has already occurred. Before empirically analyzing this potential moderating role, we briefly discuss how litigation impacts firm value and financial performance.

III.3 The Impact of Litigation on Firm Value

Litigation is likely to have significant negative impacts on firm value since lawsuits usually incur costs related to large settlement payments and damages to reputation, creditworthiness, and general trust. More specifically, litigation significantly increases loans and bond yield spreads and decreases credit ratings (Arena, 2018). It further raises the likelihood of financial covenants and collateral requirements in bank loan contracts, which accords with the reputational loss resulting from litigation (Deng et al., 2014). Litigation reduces the future use of external financing, investments in capital expenditures, and research and development expenses (Arena and Julio, 2015; Autore et al., 2014). Additionally, litigation directly impacts firm value. For instance, litigation triggered by underpricing in initial public offerings (IPOs) is often settled with payments in the two-digit percentage range of the total proceeds raised by IPOs (Lowry and Shu, 2002). These amounts are often significant and potentially diminish firm success.

Previous research has supported this evidence by arguing that less litigation positively affects firm value. Smith and Stulz (1985) found that perfect market assumptions often do not

[39]For both hypotheses $H1_A$ and $H1_B$, the corresponding null hypothesis, $H1_0$, is that there is no significant relation between positive and negative CSP and the probability of facing a class action lawsuit at all.

hold and that shareholder value thus benefits from reducing risk. From a long-term investor's perspective, less litigation positively influences financial performance by stabilizing cash flow and better allocating resources to strategic initiatives and investments (Sharfman and Fernando, 2008). Overall, litigation seems to have significant negative impacts on firm value and financial performance, while its absence enables firms to use their resources both more efficiently and more productively.

III.4 The Insurance-Like Effect of Corporate Social Performance During Litigation

This section discusses how positive CSP might mitigate the negative effects of litigation on firm value. Firms can build up moral capital (i.e., "reservoir of goodwill") through CSP, which acts as insurance against negative firm value effects during a corporate crisis (Godfrey, 2005; Koh et al., 2014). This notion of an insurance-like effect of CSP is grounded in stakeholder theory, which considers the interests of all internal and external stakeholders (Freeman, 2010). Positive CSP is a strategic investment to develop firm integrity and to build up moral capital among stakeholders (McWilliams et al., 2006). Therefore, genuine investments in CSP are likely to pay off in a corporate crisis.

This insurance-like effect has already been shown in other contexts. For instance, Lins et al. (2017) found an insurance-like effect of positive CSP during the financial crisis of 2007–08. Further, high CSP has been shown to mitigate risk in corruption-related lawsuits (Hong et al., 2019) and for firms in industries with high litigation rates (Koh et al., 2014). Aouadi and Marsat (2018) found that firms facing controversies, yet having high CSP, display higher firm values. However, research is still far from conclusive as to whether CSP protects firms against losses from legal disputes. Very recently, Liu et al. (2020) found that CSP does not offer insurance-like effects to protect market value in environmental lawsuits.

We perceive a class action lawsuit as a potential corporate crisis (hence similar to a financial crisis, ESG controversy, or corruption-related lawsuit). To the best of our knowledge, no previous study has investigated whether this insurance-like effect of CSP also holds during

actual class action lawsuits. Based on the notion of a "reservoir of goodwill" and on the resulting insurance-like effect of CSP (Godfrey, 2005; Godfrey et al., 2009; Koh et al., 2014), we argue that firms with strong positive CSP experience significantly lower declines in firm value caused by actual class action lawsuits than firms with weak positive CSP. From an investor's perspective, understanding these significantly lower declines in firm value help to explain the impact of "tail" events (i.e., unlikely but highly negative returns) such as class action lawsuits. Based on the argumentation above, we formally state the hypothesis:[40]

H2: Firms with higher positive CSP (i.e., stronger ESG performance) exhibit smaller losses in terms of (cumulative) abnormal returns during class action lawsuits than firms with lower positive CSP.

III.5 Testing the Relation between CSP and Litigation Risk

This section tests the first hypothesis (H1) whether higher positive CSP and lower negative CSP are linked to a lower probability of facing litigation.

III.5.1 Measuring Corporate Social Performance

We estimated positive and negative CSP using two distinct measures for ESG performance and ESG controversies. ESG performance measures positive corporate activities along all three ESG dimensions (environmental, social, governance). This measure is designed to capture how a firm performs across ESG themes such as emissions, product innovation and responsibility, human rights, and shareholder rights. We refer to ESG performance as the ESG_P-score in our model—capturing the upside potential of positive CSP.

In contrast, ESG controversies measure the degree to which a company violates environmental, social, and governance norms. ESG controversies fundamentally differ from ESG performance because they capture the downside risk of negative CSP. We refer to ESG con-

[40]The corresponding null hypothesis, $H2_0$, is that there is no significant relation between positive CSP and abnormal returns during class action lawsuits at all.

troversies as the ESG_N-score in our model. The ESG_N-score is mainly designed to report the number of controversies a firm was involved with in the past year, and is corrected for an industry effect. Thus, it is a very simple and parsimonious measure for negative CSP. However, this ESG_N-score is advantageous compared to just counting a firm's controversies as it benchmarks that number against industry peers. Controversies are more common in some industries than in others (Francis et al., 1994). Our measure for ESG controversies accounts for the fact that merely using the number of controversies as a proxy would induce industry bias to our analysis.

III.5.2 Proxies for Litigation Risk

We used three distinct binary variables to proxy for litigation risk as each variable measures litigation risk from a slightly different angle.

The first variable, the occurrence of a class period (ClassPeriod), measured whether the defendant was involved in potentially punishable endeavors at the time. The class period in class action lawsuits is the period in which the defendant committed the alleged injury or infringement against the class. From a panel data perspective, if a class period spans several years, the variable ClassPeriod took the value of 1 in each respective year.

The second variable, the occurrence of a lawsuit filing (CaseFiled), was the most definite proxy for litigation and measures whether the defendant was charged in a court of law. If this occurred for a firm in a certain year, the variable took a value of 1 and 0 otherwise.

The third variable, the start of a class period (ClassPeriodStart), only indicates whether the potentially punishable endeavor started in that respective year. It often takes several years for plaintiffs to file a case. We included all three variables in our model for additional robustness of results.

III.5.3 Estimating the Impact of CSP on Litigation Risk

During the last 25 years, research on litigation risk has widely used Francis et al.'s (1994) (henceforth FPS) industry membership measure as a proxy. This measure of litigation risk

(henceforth FPS measure) uses membership in biotechnology, retail, and information technology industries (henceforth FPS industries) to proxy for litigation risk. Historically, these industries have the highest litigation rates (Francis et al., 1994). One stream of research has focused on what constitutes litigation risk and how it affects firm performance. Most of this research either uses some FPS industry dummy or only includes firms from FPS industries in its sample to proxy for litigation risk (e.g., Matsumoto, 2002; Johnson et al., 2001; Ajinkya et al., 2005). However, recent research has increasingly pointed out that this crude measure of litigation risk is by far not the best way to capture litigation risk.

Instead, the idea of including additional firm and stock return characteristics into models of litigation risk has evolved (e.g., Rogers and Stocken, 2005). Kim and Skinner (2012) built on these findings and showed that firm and return characteristics substantially increase the model's predictive power.[41] We therefore used the "best" model, as identified by Kim and Skinner (2012) (henceforth KS model), and adjusted it with our ESG_P-score and ESG_N-score.[42]

Adopting and adjusting the KS model using the ESG_P-score and ESG_N-score had two basic motivations: First, applying the most efficient litigation risk model available to date enabled minimizing the omitted variables bias while ensuring a large sample. Since we already included those variables that best capture litigation risk (Kim and Skinner, 2012), our model likely predicts the "true" value of ESG performance and ESG controversies to explain litigation risk. Second, including ESG performance and ESG controversies into the KS model to estimate litigation risk is a simple and parsimonious approach to examining the explanatory power of positive and negative CSP.

In addition to the FPS measure, we included control variables for firm size, sales growth, return, return skewness, return standard deviation, and stock turnover—comprehensively discussed in Kim and Skinner (2012). We lagged all variables, except FPS, by one year since we

[41] Kim and Skinner (2012) found that only using the FPS measure to proxy for litigation risk, shows a very low overall goodness of fit and predictive ability (McFadden pseudo–R^2 and Cox&Snell pseudo–R^2 of 0.01% and 0.20%, respectively). After adding lagged firm and return characteristics to their model, overall goodness of fit and predictive ability increased considerably (McFadden pseudo-R2 and Cox&Snell pseudo–R2 of 12.47% and 4.65%, respectively).

[42] Kim and Skinner (2012) tested various models with different sets of variables. In terms of model selection, we adopted their cost-benefit perspective. They found that including more variables marginally increases the goodness of fit and predictive ability. However, collecting data for ever more variables is costly for the researcher in terms of time and lower sample size. Therefore, we decided to adhere to the most cost-benefit efficient model, as suggested by Kim and Skinner (2012).

were interested in the ex-ante predictive ability of ESG_P-score and ESG_N-score. Measuring the variables in the year during the lawsuit would have included stock price movements that trigger litigation rather than ex-ante litigation risk (Kim and Skinner, 2012). Hence, lagging the variables accounted for potential reverse causality concerns. Therefore, our empirical model looks as follows:

$$\begin{aligned} \text{Prob}(Litigation = 1) = \gamma_0 &+ \gamma_1 \cdot ESG_P\text{-score}_{t-1} + \gamma_2 \cdot ESG_N\text{-score}_{t-1} \\ &+ \gamma_3 \cdot FPS_t + \gamma_4 \cdot Size_{t-1} + \gamma_5 \cdot SalesGrowth_{t-1} \\ &+ \gamma_6 \cdot Ret_{t-1} + \gamma_7 \cdot RetSkew_{t-1} \\ &+ \gamma_8 \cdot RetStdDev_{t-1} + \gamma_9 \cdot Turnover_{t-1} + \varepsilon \end{aligned} \tag{III.5.1}$$

where $\text{Prob}(Litigation = 1)$ is a binary variable that measures litigation risk; $ESG_P\text{-score}_{t-1}$ is a continuous variable that measures ESG performance; $ESG_N\text{-score}_{t-1}$ is a continuous variable that measures ESG controversies; FPS_t is a binary variable that measures industry membership in one of the FPS industries; $Size_{t-1}$ is a continuous variable that measures firm size; $SalesGrowth_{t-1}$ is a continuous variable that measures growth in sales; Ret_{t-1} is a continuous variable that measures stock returns; $RetSkew_{t-1}$ is a continuous variable that measures skewness of stock returns; $RetStdDev_{t-1}$ is a continuous variable that measures standard deviation of stock returns; and $Turnover_{t-1}$ is a continuous variable that measures stock turnover. We obtained control variable data from Compustat/CRSP and winsorized the accounting- and return-related continuous variables at the 1%-level. We checked for multicollinearity by calculating variance inflation factors for the full model. No variance inflation factor exceeded 2.75, indicating no serious multicollinearity issues. For detailed variable descriptions, please refer to Table B.1 in the appendix.

III.5.4 ESG Data

We used data from Thomson Reuters to measure positive and negative CSP. These data are designed to reflect a firm's performance on a broad variety of ESG issues. Thomson Reuters

rates firms according to 178 carefully selected ESG issues. The underlying measures are assessed according to criteria like comparability, data availability, and industry relevance. The Thomson Reuters database is continuously updated and scores are recalculated weekly. However, scores are mostly updated once a year, depending on firms' ESG disclosure.[43] Unlike ESG scores, ESG controversy scores are updated more frequently, based on the occurrence of controversies and media reports.[44] Moreover, Thomson Reuters deems the ESG controversy score to be a viable standalone measure of a firm's performance on controversies. Thomson Reuters scores ensure comparability, data availability, and industry relevance with a reasonable level of variability, thus facilitating statistical inference. Hence, these scores were appropriate for our analysis, which aimed to understand the potential of these scores in order to explain the probabilities of future litigation.

Therefore, we used the scores for ESG performance and ESG controversies for all US firms for the period 2003–2017 from Thomson Reuters. The scores for ESG performance and ESG controversies range from zero to one. For instance, a score of .95 indicates that this firm belongs to the top 5% of firms regarding ESG performance or ESG controversies. In other words, the higher the score, the better the firm performs on ESG issues and the fewer ESG controversies occurred in the past, respectively. For our panel data set, we retrieved 9683 firm-year observations.

III.5.5 Class Action Lawsuit Data

We obtained US securities class action data from the SCAC of Stanford Law School. The data is open access and available via the corresponding website.[45] These data on class action lawsuits stem directly from US court records (Liu et al., 2020; Aharony et al., 2015; Haslem, 2005). compared to newspaper sources (e.g., Flammer, 2013) or measures derived from industry

[43] We explicitly do not use ASSET4 ratings, the predecessor to Thomson Reuters ESG scores. Instead, we use the more recent Thomson Reuters ESG scores. Thomson Reuters ESG scores were spun-off into a new entity called "Refinitiv" during the conceptualization of this study. However, the methodology has not significantly changed with this spin-off. More information about the methodology can be retrieved from: https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/esg-scores-methodology.pdf

[44] Thomson Reuters also provides a combined ESG score, which merges the overall ESG score with the ESG controversy score. However, we do *not* use this combined ESG score in our study.

[45] The website can be accessed via: http://securities.stanford.edu

membership (e.g., Koh et al., 2014). Following Kim and Skinner (2012), we only collected data for lawsuit filings against listed firms in the US (i.e., mostly from the New York Stock Exchange, American Stock Exchange, or National Association of Securities Dealers Automated Quotations) and excluded filings that are related to IPO allocations, mutual funds, and analysts.

According to Kim and Skinner (2012), roughly 89% of the lawsuit filings contained in the database are related to misstated or omitted material information (i.e., violations of SEC Rule 10b-5). Most class actions (i.e., the plaintiffs) accuse firms (i.e., the defendant) of material misstatements about the business, of failing to inform in time, or of inadequate internal control (Kim and Skinner, 2012). Due to this homogeneity of the types of lawsuits, we refrain from differentiating between the allegation types in this study. The SCAC data on lawsuit filings contain information such as case filing date, the industry sector, market status, and class period. The filing date and class period constitute our three proxies for litigation risk.

After obtaining the data on class action lawsuits, we merged them with our ESG data set. This resulted in 734 firm-year observations with a class period (ClassPeriod), 303 with a lawsuit filing (CaseFiled), and 276 with a start of a class period (ClassPeriodStart) from a total of 9683 firm-year observations. The sample size was reduced to 7671 firm-year observations after removing observations with missing control variable data and after lagging variables.

III.5.6 Descriptive Statistics

Table 10 shows summary statistics for all variables from Model III.5.1. An average sample firm has a per-year probability of 7.6% for being involved in a class action lawsuit. However, the average probability of a case being filed is only 3.1% per year. This discrepancy suggests that class periods usually last two to three years before filing. While case filings are more definite, class period years are also a valid measure of litigation risk as they are an early sign of future case filings. The FPS variable indicates that roughly 25% of sample firms belong to one of FPS industries (biotechnology, information technology, retail). This picture is confirmed when reviewing firm frequencies per sector. All three FPS industries are well represented in our sample (see Table B.3 in the appendix). Consistent with Francis et al. (1994), subsample descriptive statistics show that firms from FPS industries have 27% more class period years,

37% more case filings, and 35% more class period starts (see Table B.4 in the appendix).

The sample firms have an ESG_N-score mean of 0.465 (see Table 10), indicating that they faced slightly more ESG controversies than all firms from the Thomson Reuters universe with a mean of 0.5.[46] However, the ESG_N-score median of 0.573 indicates a slightly left-skewed distribution. In fact, looking at the empirical cumulative distribution function of ESG_N-score reveals a wrinkle between about 0.3 and 0.5. This finding suggests that our sample firms tend to be located either in the lower (i.e., < 0.3) or in the upper (i.e., > 0.5) part of the ESG_N-score continuum, indicating adequate variation in our data. This variation is also supported by a comparably high difference between minimum and maximum values for ESG_N-score (0.001 respectively 0.864).

Table 10: Summary Statistics

	Mean	SD	Min	Median	Max
ClassPeriod	0.076	0.266	0.000	0.000	1.000
CaseFiled	0.031	0.173	0.000	0.000	1.000
ClassPeriodStart	0.028	0.164	0.000	0.000	1.000
ESG_P-score	0.470	0.170	0.089	0.436	0.980
ESG_N-score	0.465	0.219	0.001	0.573	0.864
FPS	0.248	0.432	0.000	0.000	1.000
Size	8.767	1.493	5.463	8.642	13.309
SalesGrowth	0.042	0.192	−0.875	0.056	0.518
Ret	0.005	0.024	−0.096	0.002	0.120
RetSkew	0.001	0.007	−0.017	0.000	0.019
RetStdDev	0.003	0.006	0.000	0.001	0.043
Turnover	0.004	0.007	0.000	0.001	0.045
BTM	0.489	0.378	−0.168	0.404	2.130
Leverage	0.922	2.005	−6.460	0.558	13.002
ROE	0.165	0.378	−1.624	0.157	2.131

Notes: This table reports time-series averages of cross-sectional mean and SD, the minimum, the 50th percentiles (median) and the maximum for all dependent, independent, and control variables. The number of observations (N) is 7671.

In absolute terms, class action lawsuits seem to have occurred more frequently during the last years of our sample period (see Table B.5 in the appendix). However, relative to the number of observations in our sample, class action lawsuits have not considerably increased during the last years of our sample period. Therefore, the larger absolute numbers of class

[46] Thomson Reuters applies an industry benchmark system for their scores. For instance, a score of 0.7 indicates that the respective firm belongs to the 30% best-performing firms. Therefore, the mean of all firm scores in the Thomson Reuters universe turns out to be 0.5.

action lawsuits in the later years of our sample can be mainly attributed to a larger sample size in those years. Ultimately, however, we cannot rule out that a strong concentration of class action lawsuits in certain years drives our results. Publicly available data from the SCAC even points to an increasing trend of class action lawsuits, when considering all class action lawsuits in the database. Hence, we included year FE in our panel Probit regression models, in order to account for a possible increasing trend and spikes in class action lawsuits.

III.5.7 Estimating Litigation Risk Using Probit Regressions

We estimated one model for each of the three dependent variables (ClassPeriod, CaseFiled, and ClassPeriodStart), including all independent and control variables. Similar to Johnson et al. (2000) and Rogers and Stocken (2005), we used a Probit model to estimate the probability of class action lawsuits. However, using a Logit model does not qualitatively change our results (see Table B.6 in the appendix). We can not reject the null hypothesis of $H1_A$ and thus find that ESG_P-score does not explain the probability of class action lawsuits (see Table 11). Litigation risk usually stems from single issues such as firms' reporting misconduct. The aggregated ESG_P-score consisting of numerous indicators is hence likely inappropriate to explain the variation in litigation risk.

However, we can reject the null hypothesis of $H1_B$ and therefore find that ESG_N-score does significantly ($p < 0.01$) explain the probability of class action lawsuits. ESG_N-score negatively correlates with litigation risk with coefficients reaching from -0.45 to -0.57. As the ESG_N-score counts controversies, single specific cases of firms' misconduct do not get lost in an aggregate of broad strategies. Our model predicts that firms with fewer prior-year ESG controversies are less likely to be sued by US class actions. From an investor's perspective, this finding shows the importance of ESG controversies in terms of a risk-reducing effect to manage idiosyncratic portfolio risk. In line with risk management theory, risk-reducing characteristics of a lower negative CSP and more symmetric information seem to be at work here.

By calculating ME, we showed how strongly the ESG_N-score affects the predicted probability of a class action lawsuit. As shown at the bottom of Table 11, the coefficients translate into ME of -0.057, -0.032, and -0.026 of ESG_N-score on ClassPeriod, CaseFiled, and

Table 11: Estimation of Litigation Risk in Probit Regressions

	Class Period	Case Filed	Class Period Start
Indep. Variables	(1)	(2)	(3)
ESG_P-score $_{t-1}$	−0.15	−0.18	−0.38
	(0.25)	(0.22)	(0.20)
ESG_N-score $_{t-1}$	−0.45**	−0.57**	−0.47**
	(0.13)	(0.16)	(0.15)
Control Variables			
FPS	0.31**	0.30**	0.25**
	(0.08)	(0.07)	(0.07)
Ret_{t-1}	0.51	−0.59	2.33
	(1.07)	(1.24)	(1.60)
RetSkew_{t-1}	2.34	−3.00	3.83
	(3.42)	(4.69)	(4.41)
RetStdDev_{t-1}	30.65**	28.69**	15.19*
	(5.29)	(5.57)	(6.34)
SalesGrowth_{t-1}	0.74**	0.81**	0.56*
	(0.21)	(0.22)	(0.25)
Size_{t-1}	0.18**	0.15**	0.09**
	(0.04)	(0.03)	(0.03)
Turnover_{t-1}	−13.53**	−5.39	−8.64
	(5.10)	(6.06)	(6.81)
Intercept	−2.88**	−3.02**	−2.66**
	(0.35)	(0.33)	(0.34)
Year FE	Yes	Yes	Yes
ME ESG_N-score	−0.057	−0.032	−0.026
ME FPS	0.039	0.017	0.014
N	7671	7671	7671
F-statistic	6.346	6.579	3.129
p-value	0.000	0.000	0.000

Notes: This table shows the results from regressing ClassPeriod, CaseFiled, and ClassPeriodStart on ESG_N-score, ESG_P-score, and several control variables. We used Probit panel regressions and adjusted the standard errors that were used to calculate the test statistics and significance levels by a two-dimensional cluster to account for possible correlation of residuals across firm and time (Cameron et al., 2011). Marginal effects (ME) for ESG_N-score and FPS are displayed at the bottom of the table. Standard errors in parentheses. * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

ClassPeriodStart, respectively. Thus, the model predicts that an average firm decreases the risk of a class period by 1.25% ($-0.057 \cdot 0.219$) if it increases the ESG_N-score by one standard deviation (i.e., 0.219, see Table 10). Considering that the probability of having a class period is 7.6% for an average firm, a decrease of 1.25% percentage points is certainly considerable. Even more so when considering that an increase in the ESG_N-score by one standard deviation is still conservative. Especially firms in the lower ESG_N-score spectrum can potentially increase their ESG_N-score to a larger extent than a one standard deviation (analogous to decreasing marginal utility). Moreover, the model predicts that the risk of case filing drops by 0.70% ($-0.032 \cdot 0.219$) if the ESG_N-score is increased by one standard deviation. This drop is substantial. Considering the total risk of a case filing is 3.1%, dropping this risk by 0.70% percentage points to 2.40% is a decrease of 22.6%.

Consistent with Kim and Skinner (2012), we further found that being a member of an FPS industry translates into a higher litigation risk between 1.4% (for ClassPeriodStart) and 3.9% (for ClassPeriod).[47] However, being part of an FPS industry is not at the discretion of managers. Hence, from a management and investment perspective, reducing this source of risk is less attractive. Managers can only pull out of the industry and thus most likely completely undermine their business model. Investors would have to exclude highly litigious industries from their investment universe. This exclusion would most likely limit diversification potential in investment portfolios. The decrease in litigation risk through an increased ESG_N-score, however, has far more important implications. For instance, investors could consider data on ESG controversies in their investment decisions without excluding whole industries from their investment universe.

We also report the McFadden Pseudo-R^2 and the adjusted McFadden Pseudo-R^2 as measures of the goodness of fit of the Probit models. Comparing the McFadden Pseudo-R^2 statistics in Table 11 with the one from the "baseline" model of Kim and Skinner (2012) (see Table B.7 in the appendix) shows that including the ESG variables increases overall goodness of fit of the model (on an adjusted and non-adjusted basis). In particular, the adj. McFadden Pseudo-R^2 in our setting improved by 0.32 (relative increase of 6.4%), 0.53 (10.9%), and 0.37 (28.7%) for

[47] The coefficients of the other control variables (Ret, RetSkew, RetStdDev, SalesGrowth, Size, and Turnover) are also in line with Kim and Skinner (2012).

the ClassPeriod, CaseFiled, and ClassPeriodStart specifications, respectively.

Although we estimated the Probit models with time-fixed effects and additional ESG variables, our results might not fully capture the influence of other variables such as ownership structure. We try to address this limitation by the most cost-benefit-efficient model of Kim and Skinner (2012) and only tested for the marginal effects of ESG controversies and ESG performance. We followed extensive previous research that has already examined a large set of potentially relevant explanatory variables for litigation risk (Kim and Skinner, 2012; Johnson et al., 2000; Brown et al., 2005; Rogers and Stocken, 2005). The latter three studies found strong evidence for market- and return-variables to play a vital role in explaining litigation risk—above the FPS variable from Francis et al. (1994). Moreover, Kim and Skinner (2012) investigated additional variables such as R&D, Goodwill, PP&E, Insider Trading, and Institutional Ownership (cf., "Model 5" in (Kim and Skinner, 2012, p.304)). While they showed that they can slightly increase the explanatory power in Model 5, they also stressed that this increase comes at great "cost" for the researcher (e.g., in terms of a smaller sample size and larger amounts of data to collect). As far as institutional ownership is concerned, Kim and Skinner (2012) indeed showed a positive correlation with litigation risk. However, to avoid any concerns about the arbitrariness of our model, we ultimately decided to adhere to the most cost-benefit-efficient model of Kim and Skinner (2012) (cf. Model 3b on page 304).

III.6 Testing the Insurance-Like Effect in Class Action Lawsuits

This section examines whether positive CSP builds up moral capital that provides insurance-like effects in corporate crises by mitigating negative market impacts of class action lawsuits on firm value. We estimated the market impact of class action lawsuit filings on firm value by considering the daily abnormal firm performance around a filing in an event study setting. We measured abnormal performance as the difference between actual firm return and a benchmark return. To test for statistical inference, we calculated cumulative abnormal returns (CARs) based on the daily ARs around the event date (filing date).

We tested whether positive CSP offers a "reservoir of goodwill" and insurance-like ef-

fects around a class action lawsuit filing. As shown, a lawsuit filing can have severe financial and reputational consequences and thus serves as a negative event with unexpected outcome. Based on the notion of an insurance-like effect of CSP, firms with higher positive CSP should benefit from their higher moral capital in corporate crises. For example, firms with higher positive CSP should exhibit lower negative CARs than firms with lower positive CSP around class action lawsuit filings. However, the different dimensions of positive CSP (environmental, social, and governance) might have distinct effects on a firm's CAR (Humphrey et al., 2012, and references therein). Therefore, we separately test these effects in a cross-sectional regression design, controlling for various firm characteristics.

III.6.1 Market-Based Estimation of the Insurance-Like Effect

We first applied a market-based event study to analyze the relationship between CSP and the market returns of firms affected by class action lawsuits. We set the event day on the filing date and calculated ARs for the event window surrounding the event day $[-10,+10]$ (see, e.g., Krüger, 2015). We used a pre-event period of 10 days to prevent underestimating the loss in firm value due to anticipation effects prior to lawsuit filings (Gande and Lewis, 2009).

Using a market-based estimation, we first calculated expected returns (ERs) for firm j and day t ($ER_{j,t}$) by using Fama and French's (1993) three-factor model:

$$ER_{j,t} = \alpha + R_f + \beta_m \cdot (R_m - R_f) + \beta_s \cdot SMB + \beta_v \cdot HML \tag{III.6.1}$$

where $ER_{j,t}$ is the expected return of firm j on day t; R_f is the risk-free return rate; R_m is the return on a value-weighted market portfolio; SMB measures excess return of small caps over big caps; HML measures excess return of stocks with a high book-to-market ratio over stocks with a low book-to-market ratio; α is unexplained by the model. By applying Fama and French's (1993) multifactor risk model, we captured possible effects on abnormal returns stemming from differences in firm fundamentals (e.g., firm size and book-to-market value).

From the actual returns (Rs) of each firm j and day t ($R_{j,t}$), we subtracted the expected re-

turns ($ER_{j,t}$) to calculate abnormal returns ($AR_{j,t}$) for each day t in the event window $[-10,+10]$ surrounding each event date:

$$AR_{j,t} = R_{j,t} - ER_{j,t} \tag{III.6.2}$$

To calculate a firm's cumulative abnormal return ($CAR_{j,\tau}$), we simply summed each firm j's ARs for the period $[-10,\tau]$:

$$CAR_{j,\tau} = \sum_{t=-10}^{\tau} AR_{j,t} \tag{III.6.3}$$

III.6.2 Subsample on Class Action Lawsuit Filings

From our base panel sample of 9683 firm-year observations, 303 observations included a class action lawsuit (i.e., the number of total lawsuit filings in the sample, represented by the variable CaseFiled). We followed previous research by constituting each lawsuit filing as a firm-specific event (Gande and Lewis, 2009; Liu et al., 2020; Karpoff et al., 2008) and assigned the firms' respective ESG_P-score.

We split our sample of 303 class action lawsuits into tertiles and ranked class action lawsuits by the firms' ESG_P-score. The top tertile only contained lawsuits from firms whose ESG_P-score ranked in the upper tertile. The mid tertile included lawsuits from firms with an ESG_P-score between the upper and lower tertile. The bottom tertile contained lawsuits from firms with an ESG_P-score in the lower tertile. This approach ensured adequate within-sample variability.

III.6.3 Evidence on the Insurance-Like Effect of Corporate Social Performance

Figure 3 shows the mean CARs for the subsamples and the event window of $[-10,+10]$. Most noticeably, top ESG firms experience considerably lower negative CARs than mid or bottom

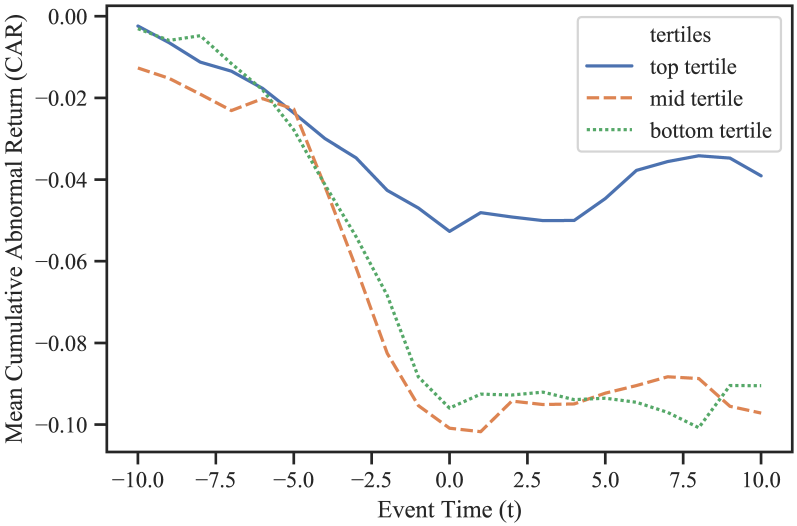
ESG firms. While CARs of top ESG firms reach their lowest point at about -5% on the event day, mid and bottom ESG firms exhibit their lowest CARs at about -11% and -10%, respectively. Mean CARs at the end of the event window are -3.91% for top ESG firms, -10.18% for mid ESG firms, and -9.05% for bottom ESG firms. For detailed statistics on ARs and CARs, please see Table B.8 in the appendix. We applied two-sample t-tests to compare CARs of the mid and bottom ESG tertile with the top ESG tertile firms. The two differences, -6.27% and -5.14%, respectively, are both significantly smaller than zero at a 5% significance level.[48] Therefore, we can reject the null hypothesis of H2.

Table B.8 and Figure B.1 in the appendix also shows that the pre-event negative ARs of top ESG firms are substantially smaller compared to ARs of mid and bottom ESG firms. While daily mean ARs for top ESG firms are only in the magnitude of -0.5%, mid and bottom ESG firms experience mean ARs of approximately -1.5%. Moreover, abnormal returns of top ESG firms seem to be recovering more strongly from their lowest point at event day. The pattern that the market reaction accrues prior to the event date also becomes clear in Figure 3 and is consistent with the anticipation effect documented in Gande and Lewis (2009). Hence, higher positive CSP does lead to smaller negative abnormal returns during litigation, supporting the notion of insurance-like effects and a "reservoir of goodwill" to firms facing a corporate crisis.

We explain the finding that firms from the mid and bottom tertile show no substantial difference in (cumulative) abnormal returns by the non-linear relationship between CSP and financial performance (see, e.g., Barnett and Salomon, 2006; Khan et al., 2016; Hong and Kacperczyk, 2009). Particularly firms with a mediocre ESG_P-score fail to implement high-level and genuine ESG solutions but still occur expenses for their ESG investments. Thus, their partial ESG commitment does not lead to insurance-like protection. Consequently, their ESG investments are inefficient in terms of Friedman (1970) and Goss and Roberts (2011).

[48] The mean difference of either groups is calculated as the mean CAR at the end of the event window. For the mid-minus-top tertile portfolio, we calculated $-10.18\% + 3.91\% = -6.27\%$, for the bottom-minus-top tertile portfolio, we calculated $-9.05\% + 3.91\% = -5.14\%$.

Figure 3: Mean Cumulative Abnormal Returns – Market Approach



Notes: This figure shows mean CARs across all 303 litigation cases in the event window from −10 days to +10 days relative to the respective event date. We split the sample into tertiles by event-year ESG_P-score.

III.6.4 Robustness Checks

Our robustness checks confirm our main result: an insurance-like effect of CSP during litigation. First, in addition to Fama and French's (1993) three-factor model, we alternatively used a market-adjusted model from CRSP that defines AR in excess of a value-weighted market return. We also used Carhart's (1997) four-factor model to calculate the expected and abnormal returns. None of these alternative models qualitatively changed our main result (see Table B.9 in the appendix).

Second, we varied the estimation window to calculate the expected returns from 250 days to 500 days and increased the gap between the estimation and the event window from 50 days to 250 days. This increased gap prevents a bias in our expected returns caused by anticipation effects prior to the lawsuit filing. We also shortened the event window to 11 days ($[-5, +5]$). None of these alternative specifications qualitatively changed our main results (see Figure B.2 and Figure B.3 in the appendix).

Third, we used alternative cutoff points to assign the firms into four instead of three groups based on their ESG score (i.e., we used quarterlies instead of tertiles). Our main result also holds under these alternative subsamples (see Figure B.4 in the appendix).

Fourth, we ran a "placebo" event study by shifting the event dates one year into the future. On this artificial event day, no significant AR should exist, and the returns of firms with different ESG scores should not differ substantially. Daily ARs are fairly low (between 0.5% and 1%), and mostly statistically insignificant. CARs range from 1.4% to 3.1% and are also statistically insignificant (see Figure B.5 in the appendix). Alleviating concerns about firm selection bias, the results of this "placebo" event study further corroborate our main result—lower negative ARs and CARs for firms with high CSP on the days surrounding a class action lawsuit filing.

Fifth, we winsorized the CARs from our market-based assessment on a 2.5%-level to mitigate the effects of outliers on the results. Winsorization did not substantially change our main result (see Figure B.6 in the appendix).

III.6.5 Cross-Sectional Analysis After Applying Propensity Score Matching

Although our market-based event study revealed a clear pattern—an insurance-like effect of high CSP during litigation—unobserved variables and other factors than CSP may influence the relationship. For instance, firms from the Communication Services sector report by far the most negative CARs around class action lawsuit filings. The Consumer Staples sector even reports slightly positive CARs during litigation. Overall, differences between sectors can be quite substantial (see Table B.10 in the appendix).

Therefore, we adopted a propensity score matching (PSM) approach to rule out that other factors (e.g., size or profitability) drive our result, i.e., lower negative CARs of top ESG firms around class action lawsuit filings. Our PSM approach also controls for this industry effect (as described) by matching control firms of the same industry group as the treated firms. Propensity scores were calculated for the control firms using various firm characteristics (industry group, ESG_P-score, book-to-market ratio, leverage, return on equity (ROE), and size). We derived

the control sample of 6097 firm-year observations from our main sample. Firms in the control sample faced no class action lawsuit filing during the whole sample period. Nevertheless, an ESG score was available. Thus, we excluded prior, contemporary, and later confounding class actions from the control sample that may influence the returns of control firms.

The treated sample contained firm years with a class action lawsuit filing. Our original subsample of 303 filings was reduced to 229 due to missing data on firm characteristics and the fact that we matched firms in the year previous to the filing. Finally, for each of the 229 treated firm-year observations, we obtained a control firm-year observation with firm characteristics being as close as possible to the treated firm (i.e., with the highest possible propensity score). This procedure yielded 229 pairs of treated and control firms, which shared the same firm characteristics one year prior to litigation. However, only the treated firm experienced a class action lawsuit filing in the year post-matching.

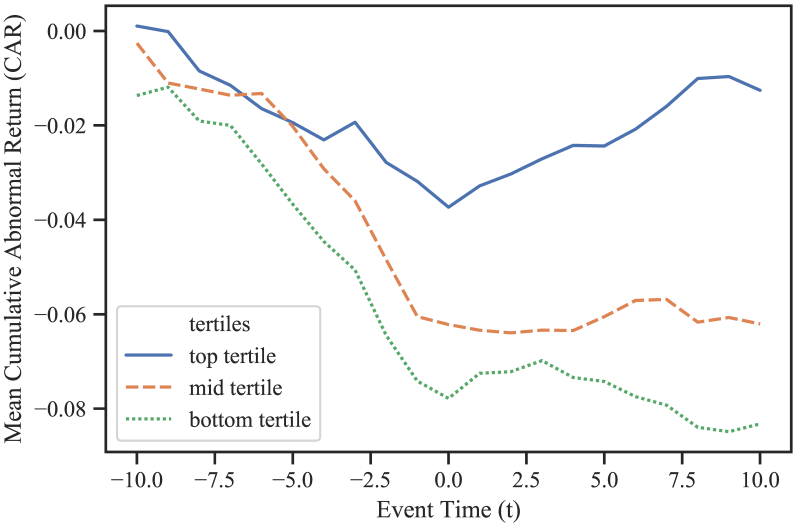
To calculate abnormal returns of treated firms ($AR_{j,t}$), we replaced the expected return of treated firms ($ER_{j,t}$) in Equation III.6.2 with the matched control firm's actual return ($R_{m,t}$). Hence, Equation III.6.2 was updated as follows:

$$AR_{j,t} = R_{j,t} - R_{m,t} \tag{III.6.4}$$

The treated firm's cumulative abnormal return ($CAR_{j,\tau}$) was calculated the same way as in the market-based estimation (cf., Equation III.6.3). We matched the respective ESG_P-score of the treated firm to the cross-section of $CAR_{j,\tau}$. We winsorized CARs at a 2.5%-level to control for a possible outlier effect due to extreme events.

Our results for CARs of treated firms accord with our previous findings from the market-based assessment of CARs. The losses of top ESG firms are only around half as high as those of bottom or mid ESG firms (see Figure 4). Thus, applying PSM confirms our market-based result of top ESG firms experiencing significantly less negative CARs around class action lawsuit filings. In absolute terms, however, negative cumulative abnormal returns are slightly lower for all firms. Interestingly, though, top ESG firms' CARs also recover better after the filing date ($[0,+10]$).

Figure 4: Mean Cumulative Abnormal Returns – Propensity Score Matching



Notes: This graph shows the mean CARs of the 239 treated firms over their matched control firms from the propensity score matching. Mean CARs are calculated for the event window from −10 days to +10 days relative to the respective event date of the treated firm.

We further conducted a cross-sectional analysis to test whether the level of the ESG_P-score impacted $CAR_{j,\tau}$. We hence explained the variation in the $CAR_{j,10}$ by the ESG_P-score and additional firm characteristics such as leverage (Bhandari, 1988), profitability as measured by ROE (Fama and French, 2015), and size (Schwert, 1983) in a cross-sectional regression. All independent variables were lagged by one year to alleviate concerns about reverse causality. We applied a percentile ranking to ESG_P-score, ENV_P-score, SOC_P-score, and GOV_P-score to account for a possible measurement bias in the ESG data (following Utz, 2018). Thus, we captured the impact of potential erroneous differences in measuring these variables on the results. Using percentile ranking for ESG_P-score, enabled maintaining the same order between the variable observations. We used bootstrapped sampling to calculate the test statistics due to our comparably small sample size (N=229).

The ESG_P-score is positively related to a firm's CAR around class action lawsuit filings (see Table 12). We additionally consider the standalone insurance-like effect of the environmental (ENV_P-score), social (SOC_P-score), and governance (GOV_P-score) dimensions of CSP. On the one hand, the social pillar of ESG seems to drive the positive relationship between CSP and CAR. On the other, the governance and environmental pillars are insignificant. This lack of significance is supported by the idea that the single dimensions of ESG differ too strongly (Humphrey et al., 2012, and references therein). After all, our finding that higher positive CSP is linked to lower negative CARs around class action lawsuits is corroborated—at least as far as the social pillar of ESG is concerned.

Table 12: Cross-Sectional Regression of $CAR_{t=10}$

	Dependent variable: $CAR_{t=10}$			
Independent Variables	ESG_P-score	ENV_P-score	SOC_P-score	GOV_P-score
Positive CSP	5.766*	4.454	6.537*	1.587
	(2.633)	(2.565)	(2.993)	(2.985)
Control Variables				
Leverage	−0.357	−0.345	−0.343	−0.372
	(0.214)	(0.217)	(0.202)	(0.234)
ROE	0.826	0.841	0.776	0.866
	(0.674)	(0.715)	(0.835)	(0.728)
Size	0.744	0.890*	0.696	1.099*
	(0.450)	(0.409)	(0.449)	(0.448)
Intercept	−14.323**	−15.067**	−14.277**	−15.540**
	(4.069)	(4.107)	(3.826)	(4.405)
N	229	229	229	229
Adj. R2	0.06	0.05	0.06	0.05
F-statistic	18.48	19.24	20.59	10.36
p-value	0.001	0.001	0.000	0.035

Notes: This table shows the results from cross-sectional OLS regressions of CAR_{t-10} from the propensity score matching on ESG_P-score (incl. sub-scores) and control variables using robust and bootstrapped standard errors. ESG_P-score and its sub-scores (ENV_P-score, SOC_P-score, and GOV_P-score) denote the variable used for Positive CSP. $CAR_{t=10}$ are in percent. ESG_P-score and its sub-scores were recalculated as within-sample percentile ranks to account for possible measurement bias. Standard errors in parentheses. * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

III.7 Implementation of the Findings in a Trading Strategy

In this section, we synthesize the findings of a lower predicted litigation risk due to fewer ESG controversies and insurance-like protection through CSP during litigation. Therefore, we tested whether implementing our findings in a trading strategy yields positive alpha. Our results show that portfolios comprising stocks with low predicted litigation risk show significantly higher alphas than portfolios of stocks with high exposure to litigation. Moreover, suppose an investment manager has to select stocks with high litigation risk (for instance, due to investment style or restrictions). In that case, our results suggest that high CSP mitigates negative performance.

III.7.1 Factor Portfolios and Abnormal Risk-Adjusted Returns

In a first step, we predicted probabilities of class action lawsuits (=litigation risk) for each sample firm-year between 2014 to 2017. To calculate the probabilities for 2014, we used sample data from the years 2003 to 2012 to estimate the parameters γ of our Probit model (based on Equation III.5.1). Subsequently, we used these parameters and data from 2013 to calculate the out-of-sample probabilities for 2014. This calculation was repeated for each year while the time frame to estimate the model parameters was extended by one year with each iteration (e.g., 2003 to 2013 for the probabilities of 2015).

We checked the out-of-sample accuracy of our predictions by applying correlation tests and *t*-tests. For the correlation tests, we calculated the Pearson's correlation between the predicted probability for litigation and the actual class action lawsuits in each year for each firm. The correlations are positive and significant (at the 1% level) for all three measures of litigation. Thus, firms with higher predicted probabilities for litigation also experience more actual litigation cases in our sample. This result is confirmed when conducting two sample *t*-tests on the predicted probabilities for litigation of firm-year observations with and without actual litigation cases. Our Probit prediction model estimated significantly higher probabilities for firms that experienced an actual out-of-sample litigation case than for firms that did not face litigation (4.7% vs. 3.3%, respectively). The difference is significant at the 1% level.

In a second step, we created four single-sorted long-only factor portfolios: one containing the 20% lowest litigation risk firms ("Low Lit"), one containing the 20% highest litigation risk firms ("High Lit"), one containing the 33% lowest prior-year ESG_P-score firms ("Low ESG_P"), and one containing the 33% highest prior-year ESG_P-score firms ("High ESG_P"). In accordance with our previous results on an insurance-like effect of CSP, we also built four double-sorted lony-only factor portfolios containing the intersection of the 20% lowest litigation risk firms and the 33% highest prior-year ESG_P-score firms ("Low Lit-High ESG_P"), the 20% highest litigation risk firms and the 33% lowest prior-year ESG_P-score firms ("High Lit-Low ESG_P"), the 20% lowest litigation risk firms and the 33% lowest prior-year ESG_P-score firms ("Low Lit-Low ESG_P"), and the 20% highest litigation risk firms and the 33% highest prior-year ESG_P-score firms ("High Lit-High ESG_P"). The holding period for the portfolios

was one year.

In a third step, we retrieved the returns of the portfolio constituents from Compustat/CRSP and calculated equally-weighted and value-weighted monthly portfolio returns. Based on these portfolio returns, we estimated monthly portfolio alphas using the Fama-French 5-factor plus Momentum (FF5+MOM) model. Moreover, we swapped market returns with the returns from the STOXX North America Industry Neutral ESG USD PRICE Index as our proxy for ESG market-wide returns. Monthly portfolio alphas are *before* transaction costs.

The first set of columns of Table 13 displays the results for the four single-sorted portfolios. The equally-weighted "Low Lit" and value-weighted "Low Lit" showed significant monthly alphas of 0.75% and 0.99%, respectively. The alpha for the equally-weighted "High Lit" is negative at -0.44% and insignificant, while that for the value-weighted "High Lit" is positive at 0.44% and significant. This indicates that particularly small firms with high litigation risk exhibit negative abnormal returns, as their weight in an equally-weighted portfolio is higher than in a value weighted portfolio. Both the equally- and value-weighted High ESG_P portfolios showed monthly alphas of 0.25% and 0.53%, respectively. Moreover, finding alphas of 0.25% and 0.73% for "Low ESG_P" is consistent with the idea that investors require higher compensation for the increased idiosyncratic risk inherent to a low ESG (e.g., Utz, 2018).

The second set of columns of Table 13 reports the results for the four long-only double-sorted portfolios ("Low Lit-High ESG_P," "High Lit-Low ESG_P," "Low Lit-Low ESG_P," and "High Lit-High ESG_P"). Both double-sorted low litigation portfolios ("Low Lit-High ESG_P" and "Low Lit-Low ESG_P") yielded similar positive and significant alphas: 0.79% and 0.74% for the equally-weighted portfolios and 0.82% and 1.14%for the value-weighted portfolios. The value-weighted "High Lit-High ESG_P" generated a significant positive alpha of 0.43%, while the alpha is insignificantly different from zero in the equally-weighted setting. The alpha of the equally-weighted "High Lit-Low ESG_P" was significantly negative.

These empirical findings reveal practical implications for investment managers. A portfolio of stocks with low estimated litigation risk performs significantly better than a portfolio composed of stocks with high litigation risk. The alpha of this long-short zero-investment portfolio as the difference of "Low Lit" and "High Lit" is 1.17% (0.53%) and is significant at a

1% (5%) level in the equal-weighted (value-weighted) setting. For the low-litigation environment, we detect no insurance-like effect of high ESG_P. However, if an investment manager is virtually unable to focus on low litigation stocks, high CSP provides an insurance-like. To test the insurance-like effect in our investment strategy, we calculated a long-short zero-investment portfolio as the difference between "High Lit-High ESG_P" and "High Lit-Low ESG_P." This portfolio yielded a positive and significant alpha of 0.91% in the equally-weighted setting. In the value-weighted setting, the alpha was also positive but insignificant (unreported results). To gauge the robustness of these results, we used Fama-French 3-factor and CAPM models and varied the cutoffs for the single- and double-sorted portfolios. The generated results are similar to the ones of our main analysis.

Table 13: Long-only Trading Strategies - Monthly Fama-French 5-Factor + Momentum Alphas

	Single-Sorted				Double-Sorted			
	Low Lit	High Lit	High ESG$_P$	Low ESG$_P$	Low Lit-High ESG$_P$	Low Lit-Low ESG$_P$	High Lit-High ESG$_P$	High Lit-Low ESG$_P$
Av. #Stocks	170	170	423	423	60	109	118	52
Panel A: Equally-Weighted								
Alpha	0.0075**	−0.0044*	0.0022*	0.0030*	0.0073**	0.0085**	−0.0013	−0.1060**
Adj. R2	0.8989	0.8790	0.9493	0.9273	0.8593	0.8777	0.8642	0.8165
Av. Exp. Ret.	0.0131	0.0018	0.0089	0.0090	0.0128	0.0138	0.0054	−0.0047
Turnover	130%	108%	65%	104%	151%	151%	86%	158%
Breakeven costs	0.58%	−0.41%	0.34%	0.29%	0.48%	0.56%	−0.15%	−6.70%
Panel B: Value-Weighted								
Alpha	0.0099**	0.0044**	0.0053**	0.0071**	0.0071**	0.0127**	0.0044**	0.0031
Adj. R2	0.8786	0.9282	0.9632	0.8684	0.8900	0.7905	0.9318	0.5056
Av. Exp. Ret.	0.0151	0.0123	0.0131	0.0133	0.0124	0.0175	0.0126	0.0929
Turnover	133%	45%	23%	81%	160%	149%	43%	111%
Breakeven costs	0.75%	0.98%	2.31%	0.88%	0.45%	0.85%	1.03%	0.28%

Notes: We used Fama-French 5-Factor + Momentum models to calculate monthly alphas. However, we swapped Fama-French market returns with the returns from the STOXX North America Industry Neutral ESG USD PRICE Index. Panel A reports monthly alphas of the equally-weighted single-sorted portfolios ("Low Lit," "High Lit," "High ESG$_P$," and "Low ESG$_P$") and double-sorted portfolios ("Low Lit-High ESG$_P$," "Low Lit-Low ESG$_P$," "High Lit-High ESG$_P$," and "High Lit-Low ESG$_P$"). We used 20% cutoffs for the litigation probability and 50% for *ESG$_P$*-score to build the portfolios. Panel B contains the monthly alpha of the value-weighted single-sorted and double-sorted portfolios. The backtesting time frame was 2014 to 2017 (N=48). The single-sorted portfolios contained 170 stocks (Lit) and 423 stocks (ESG$_P$) on average in each year. The double-sorted portfolios "Low Lit-High ESG$_P$," "Low Lit-Low ESG$_P$," "High Lit-High ESG$_P$," and "High Lit-Low ESG$_P$" contained 60, 109, 118, and 52 stocks on average in each year, respectively. To put the alphas into perspective, average monthly expected portfolio returns are displayed. We also report annual turnover and breakeven transaction costs in per cent. * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

III.7.2 Transaction Costs of Factor Portfolios

The previous results show that allocating to low litigation and high ESG_P-score portfolios leads to positive abnormal risk-adjusted returns. Since these abnormal risk-adjusted returns do neglect transaction costs, we calculated turnover and breakeven transaction costs for the single- and double-sorted portfolios. We followed Houweling and Van Zundert (2017) and defined the breakeven transaction costs as the costs that lower a portfolio's abnormal risk-adjusted return to zero. Our measure for the abnormal risk-adjusted return is the FF5+MOM alpha, i.e., the values labeled as alpha in Table 13. To assess the amount of trading required for each strategy, we calculated the turnover for each portfolio as the percentage of the wealth of the whole portfolio that is traded from month t to month $t+1$. That means, the turnover is determined as the sum of all weight increments across the portfolio constituents.

Table 13 reports the turnover results. The annualized turnover rates of the single-sorted portfolios range from 23% to 133%, while those for the double-sorted portfolios range from 43% to 160%. These substantial turnover rates indicate that investing in the suggested portfolios comes at a cost. The turnover rates stem from new stocks entering or old stocks leave the top or bottom portfolios due to changes in predicted litigation risk and ESG_P-score. The constituents of the "High ESG_P" portfolio appear to be comparably stable (turnover rates of 65% and 23%), while the constituents of the "Low Lit" portfolio strongly differ across years (turnover rates of 130% and 133%).[49] The double-sorted portfolios show even higher turnover rates, which is due to the substantially lower number of stocks in these portfolios. The single-sorted portfolios have, on average, 170 stocks for the Lit-portfolios and 282 stocks for the ESG_P-portfolios, while the double-sorted portfolios contain 33 ("Low Lit-High ESG_P"), 75 ("Low Lit-Low ESG_P"), 96 ("High Lit-High ESG_P"), and 34 ("High Lit-Low ESG_P") stocks on average in each year.

Moreover, we calculated the breakeven transaction costs of each portfolio as its alpha divided by its turnover (see Table 13). The breakeven transaction costs for single-sorted portfolios range from 0.29% to 2.31%, except for the "High Lit" portfolio that generates a negative FF5+MOM alpha. The breakeven transaction costs for the value-weighted portfolios are slightly higher than for the equally-weighted portfolios because they have higher FF5+MOM al-

[49] This result is not due to the different cutoffs, because it remains if we choose the same cutoffs for Lit and ESG_P.

phas and lower turnover rates. The double-sorted portfolios show a broader range of breakeven transaction costs. However, the "Low Lit" portfolio exhibits breakeven transaction costs of 0.48%. To put these figures into perspective, we compare them with actual transaction cost estimates. Frazzini et al. (2014) estimated that the average trading cost for a large institutional investor managing quantitative strategies in the large-cap developed-markets universe was less than 0.2% of the trading value between 1998 and 2013. These transaction costs are well below the breakeven transaction costs that we reported for our portfolios in Table 13—at least for the majority of the portfolios. Consequently, our portfolios can sustain the actual observed transaction costs and thus generate positive after-cost FF5+MOM alphas.

III.8 Concluding Remarks and Practical Implications

We have investigated whether ESG controversies in the past explain future litigation risk and whether ESG performance positively moderates the negative effects of litigation on firm value. Unlike previous studies, we have focused on the mere number of past ESG controversies to account for a firm's negative CSP. We find that fewer ESG controversies (i.e., better ESG_N-score) are linked to lower litigation risk. In particular, improving the ESG_N-score of an average firm by one standard deviation reduces the predicted probability of a class action lawsuit by 22.6%. Moreover, we show that higher ESG performance (i.e., better ESG_P-score) shields against negative events by effectively lowering the negative impacts of litigation on firm value. Our results show that an average sample firm with low ESG_P-score facing a class action lawsuit suffers an excess loss in market value of about US $1.14bn (5.14% of US $22.1bn) compared to a firm with high ESG_P-score.

From a portfolio management perspective, litigation risk is one component of a firm's risk exposure. The rationale behind our empirical finding might be the following. Although, in theory, diversification eliminates idiosyncratic risk, empirical results show that idiosyncratic risk and expected returns are positively correlated (Fu, 2009; Merton, 2006). In particular, lower litigation risk exposure results in a lower required return, leading to lower cost of capital, which allows firms to invest financial resources more efficiently. Lower cost of capital also implies higher firm value, *ceteris paribus*.

The insurance-like effect of CSP also has implications for portfolio management. Investors might include aspects of CSP in their decision-making to mitigate the financial constraints during class action lawsuits. Khan (2019) found evidence for a signaling value of ESG metrics by showing that they explain stock returns. Our research suggests that ESG metrics in particular convey information about a firm's risk exposure, which in turn influences stock returns. Moreover, high CSP might be a tool for retaining firm value during a corporate crisis and preventing firms from financial distress. Distressed firms tend toward fire sale assets at lower prices in crises (Ang and Mauck, 2011) and thus may jeopardize their long-term prospect. Hence, high CSP could be considered as a defense mechanism to acquisitions and low-price fire sales.

Synthesizing our findings on lower litigation risk and insurance-like effects of CSP offers implications for investment practice. Consistent with reduced information asymmetries, scores on the involvement of firms in ESG controversies provide useful signals to shareholders about litigation risk. Our results enable shareholders to infer that the parsimonious and simple driver of litigation risk—the mere number of ESG controversies—is easily adoptable in investment processes. Accordingly, we included such a measure of ESG controversies to predict litigation risk and to build portfolios of firms with the lowest litigation risk. These portfolios achieve positive alphas. We also built portfolios of the firms with the highest ESG performance to capitalize on the insurance-like effect during potential class action lawsuits. Combining the two criteria of low litigation risk and high ESG performance in long-only factor portfolios yields monthly alphas between 0.71% and 0.73%. Although these alphas are before transaction costs, we show that even after accounting for these costs the alphas remain positive. The size of breakeven transaction costs indicates that generating these positive alphas with our trading strategy is feasible. Moreover, the required data are readily available to be integrated in the investment management process.

IV Paper 3: Corporate Social Irresponsibility and Credit Risk Prediction: A Machine Learning Approach[50]

Daniel V. Fauser, Andreas Grüner

Abstract This paper examines the prediction accuracy of various machine learning (ML) algorithms for firm credit risk. It marks the first attempt to leverage data on corporate social irresponsibility (CSI) to better predict credit risk in an ML context. Even though the literature on default and credit risk is vast, the potential explanatory power of CSI for firm credit risk prediction remains unexplored. Previous research has shown that CSI may jeopardize firm survival and thus potentially comes into play in predicting credit risk. We find that prediction accuracy varies considerably between algorithms, with advanced machine learning algorithms (e.g., random forests) outperforming traditional ones (e.g., linear regression). Random forest regression achieves an out-of-sample prediction accuracy of 89.75% for adjusted R^2 due to the ability of capturing non-linearity and complex interaction effects in the data. We further show that including information on CSI in firm credit risk prediction does *not* consistently increase prediction accuracy. One possible interpretation of this result is that CSI does not (yet) seem to be systematically reflected in credit ratings, despite prior literature indicating that CSI increases credit risk. Our study contributes to improving firm credit risk predictions using a machine learning design and to exploring how CSI is reflected in credit risk ratings.

Key Words: Corporate social irresponsibility; Credit risk; Machine learning; Random forests; Regression; Classification

[50]This paper has been accepted for publication in the journal Credit and Capital Markets. The journal is ranked as *C* in VHB-JOURQUAL3. We thank the two guest editors Prof. Dr. Marco Wilkens and Prof. Dr. Christian Klein as well as an anonymous reviewer for their valuable comments and suggestions, which have significantly improved our paper.

IV.1 Introduction

This study investigates the prediction accuracy of machine learning (ML) algorithms for firm credit risk using data on corporate social irresponsibility (CSI). We aim to leverage the power of both advanced ML algorithms *and* data on irresponsible firm behavior with regard to ESG aspects (i.e., CSI) to improve the accuracy of credit risk prediction. Our study is motivated by research that points to the increasingly important role of environmental, social, and governance (ESG) aspects for firm credit risk. ESG performance has been linked to lower firm and credit risk (Stellner et al., 2015; Jo and Na, 2012; Sassen et al., 2016) and to higher credit ratings (Jiraporn et al., 2014; Kiesel and Lücke, 2019; Dorfleitner et al., 2020; Oikonomou et al., 2014).

Common to these previous studies is that they have predominantly used ESG performance data to examine firms' corporate social performance (CSP). One exception, however, is Kölbel et al.'s (2017) study on the exposure to negative ESG performance, based on the idea of corporate social irresponsibility (CSI) (Strike et al., 2006). They showed that CSI leads to higher credit risk as priced by the market through increased spreads of credit default swaps (CDS). However, it is still unclear whether CSI is also systematically reflected in credit ratings and thus informs credit risk predictions. Therefore, our study uses a comprehensive machine learning regression design to test how consistently CSI has been considered in the credit risk analyses of major credit rating agencies. It is important to note that our study differs from previous research on ESG performance and credit risk (e.g., Stellner et al., 2015; Dorfleitner et al., 2020; Sassen et al., 2016) as we focus on *negative* ESG performance as opposed to *positive* ESG performance.

Our study also has both a practical, "supply-driven" motivation and a practical, "demand-driven" motivation. The suppliers of credit ratings, so-called credit rating agencies, have recognized that ESG aspects are no longer solely a matter of philanthropy but may deeply impact firm performance and risk. The documentation by major credit rating agencies (such as S&P Global) indicates that ESG aspects are to some extent reflected in credit ratings. For example, S&P Global claims that it "has long considered Environmental, Social, and Governance (ESG) factors in its credit ratings [...], [capturing] ESG factors in many areas of [its] methodology." According to S&P Global, ESG aspects are considered in credit risk analysis along three

dimensions: business risk (e.g., competitive position), financial risk (e.g., cash flow/leverage assessment), and management and governance.[51] On November 21, 2019, S&P Global issued a press release on its most recent step toward ESG integration by acquiring RobecoSAM's ESG rating business.

As far as the "demand" side is concerned, the Principles for Responsible Investments (PRI) Association conducted an investor survey in 2017 and 2018 among credit analysts, portfolio managers, and ESG analysts about the integration of ESG aspects into credit risk analysis and ratings (PRI, 2018). The survey found that 62.9% of participants agreed that valuation models are adjusted as a result of ESG considerations in their organization. According to participants, the governance dimension of ESG most frequently impacts credit risk analysis, followed by the environmental and social dimensions. The survey identified the three most common reasons for including ESG aspects in credit risk analyses as "Risk management," "Client demand," and "Fiduciary duty." The results of this survey underpin the constantly increasing demand to integrate ESG aspects into credit risk analysis.

Regarding methodology, our decision to use ML algorithms was motivated by research in finance that has found evidence for the superior predictive power of ML algorithms compared to more traditional techniques (e.g., Barboza et al., 2017; Gu et al., 2020). ML has proven highly successful in addressing many problems ranging from early-day applications (e.g., spam detection) to more recent fields of application (e.g., financial forecasting and portfolio building). After long playing merely a minor role in finance, the potential of ML in financial prediction and modeling has started to gain momentum. Using ML in bankruptcy prediction (Barboza et al., 2017), empirical asset pricing and portfolio construction (Gu et al., 2020), stock selection (Rasekhschaffe and Jones, 2019), and sovereign credit ratings prediction (Bennell et al., 2006) are just a few examples. ML techniques are particularly powerful in solving prediction and classification tasks due to their ability to closely capture non-linearity, complex relationships, and interactions in the data. Common issues of traditional linear algorithms such as multicollinearity or low signal-to-noise ratios are not a serious concern for ML techniques, as we show below. Accordingly, our study tests a comprehensive set of ML algorithms for predicting

[51] For S&P Global's claims about ESG integration in credit risk analysis, please see: https://www.spglobal.com/ratings/en/products-benefits/products/sustainable-finance and https://www.spglobal.com/ratings/en/products-benefits/products/esg-in-credit-ratings#.

credit risk as measured by credit ratings.

We use an extensive international sample of 43,476 firm-quarter observations with a time-series of over 12 years (March 2007 to September 2019). To address the potential impact of CSI on credit risk, we used data from RepRisk, which specializes in ESG and business conduct risk research. Data on our measure for credit risk, long-term issuer credit ratings, were primarily obtained from S&P Global. We also used data from the two other major rating agencies, Moody's and Fitch, in an additional analysis. Our study uses algorithms ranging from "plain vanilla" linear regression to more sophisticated machine learning techniques such as Multilayer Perceptron regression. In line with previous research (e.g., Barboza et al., 2017), we used *sklearn's*[52] package default settings for all algorithms to ensure a high degree of replicability and applicability. We assessed prediction accuracy with the three most common measures in machine learning: R^2, adjusted R^2, and root mean squared error (RMSE). Using default settings also aims to alleviate the frequently expressed criticism that ML techniques are a "blackbox." Default hyper-parameters enable precisely replicating our research. To ensure transparency, all algorithms, hyper-parameters, and settings can be found in Table C.13 in the appendix and are also available in the *sklearn* API reference.

Our study has two main findings. First, using advanced ML algorithms considerably increases prediction accuracy for credit risk by accounting for non-linearity and complex interaction effects in the data. Random Forest regression offers the highest out-of-sample prediction accuracy of 89.75% for adjusted R^2, indicating that this algorithm reliably predicts credit risk two quarters into the future. This result is in line with previous research findings on the similar superiority of Random Forest algorithms in default prediction (Barboza et al., 2017) and time-series forecasting (Gu et al., 2020). Second, CSI does not (yet) seem to be consistently and systematically reflected in credit risk as measured by credit rating data. Incorporating information about CSI into our models does not unambiguously increase prediction accuracy for credit risk. This finding at least holds for both the data and the comprehensive set of ML algorithms used in this study. One possible interpretation is that CSI is not (yet) incorporated into the credit risk analyses of rating agencies to such a degree that this would result in considerably higher

[52] *sklearn* is the short version for *scikit learn*. The application programming interface (API) reference can be found on https://scikit-learn.org/stable/index.html

prediction accuracy, as is the case for profitability or liquidity, for example. Another interpretation is that rating agencies use a significantly different operationalization of ESG aspects, which is not available to the public.

Our study makes several contributions to the literature and has several implications for market participants. First, to our best knowledge, our study is the first of its kind to synthesize credit risk prediction with state-of-the-art machine learning algorithms using credit rating data. In a credit risk context, our results support the findings of previous research, among others, that advanced ML algorithms are superior to traditional algorithms, particularly in financial prediction tasks (e.g., Barboza et al., 2017; Khandani et al., 2010; Gu et al., 2020). Our finding of high prediction accuracy for advanced ML algorithms is based on high out-of-sample adjusted R^2 and low RMSE, which are robust compared to alternative specifications of our ML models. High prediction accuracy allows market participants to more reliably assess future credit risk. Our trained models are ready to use to predict the credit risk of any firm for which relevant data are available to the predictors used in our study. Our result points to the growing importance of machine learning models in the paradigmatic fintech industry. However, our predictive method fundamentally differs from methods of statistical inference. In particular, machine learning models do not tell us anything about the underlying mechanisms (economic and theoretical) between the predictor and the target variables (e.g., the reasons for a certain relationship between predictors and targets).[53]

Second, to our best knowledge, our study is the first of its kind to investigate whether including CSI into ML models increases the prediction accuracy of credit risk. In doing so, we also test the extent to which CSI is reflected in the credit risk analyses of rating agencies. Even though previous research has found a positive relationship between *positive* ESG performance and credit risk (e.g., Dorfleitner et al., 2020; Stellner et al., 2015; Sassen et al., 2016), we explicitly explore *negative* ESG performance in a novel machine learning design. Previous research has shown that market participants account for CSI in evaluating credit risk, as evidenced by increased spreads of CDS (Kölbel et al., 2017). However, using credit ratings as a proxy for credit risk, our results indicate that higher CSI is not (yet) reflected to a larger extent in the credit risk

[53] In this paper, we use the terms "predictor" and "target," which are more common in a machine learning context, yet roughly similar to the terms "independent variable" and "dependent variable," respectively.

analyses of major rating agencies. Therefore, market participants should be aware that common credit ratings might not entirely reflect the environmental and societal business conduct risks increasingly confronting firms. Examples of increased business conduct risk include weaknesses in corporate governance, low product safety, or compromised employee well-being. All of these factors may entail substantial future expenses (e.g., compensation payments) and reputational damages.

IV.2 Literature Review and Framework

This section considers related ML applications and their potentially superior explanatory power in credit risk prediction. It also discusses the empirical literature on the relationship between ESG aspects and credit risk. Based on this review, we investigate whether including CSI in ML algorithms improves credit risk predictions.

IV.2.1 Machine Learning Applications

The application spectrum of ML algorithms is vast and growing. ML applications range from cancer prognosis and prediction (Kourou et al., 2015), genetics and genomics (Zou et al., 2019), and healthcare (Jiang et al., 2017) to automated text classification (Sebastiani, 2002). ML algorithms have also been shown to better solve financial prediction and classification problems than traditional financial models. Therefore, ML algorithms potentially add value to many finance-related fields of application. For instance, machine learning can improve forecasting consumer credit risk and thus cuts costs from credit losses (Khandani et al., 2010). Machine learning may also help to solve financial prediction problems such as security pricing, portfolio construction, financial time series prediction, or risk management (Gu et al., 2020; Rasekhschaffe and Jones, 2019). Additional ML applications include sovereign credit rating prediction (Bennell et al., 2006), forecasting interest rates, corporate bond ratings, loan approvals, and identifying suspicious transactions (Bose and Mahapatra, 2001).

Most noticeably, predicting firm default has gained quite some momentum over the last 25 years with ever-evolving ML algorithms and exponentially increasing computing power and

data availability. Early studies on corporate default using "intelligent" techniques such as neural networks include Altman et al. (1994) and Boritz and Kennedy (1995). Most of these studies offered partially inconclusive results, possibly due to the comparably low computational power, data availability, or pure immaturity of the subject and algorithm implementations.

In a more recent study, Barboza et al. (2017) compared a wide range of more advanced ML algorithms to traditional ones of default prediction, such as logistic regression. Default prediction is a typical classification problem (i.e., 1 for default, 0 otherwise). Using US data, they showed that the majority of more advanced ML algorithms are significantly superior in default prediction accuracy than traditional techniques. The authors were also able to increase the total estimated prediction accuracy by about 11 percentage points using advanced ML algorithms (e.g., random forests).[54] This increased prediction accuracy is very likely due to ML algorithms' special treatment of non-linearity, complex interaction effects, and lower sensitivity to multi-collinearity. Building on these findings, we expect that ML techniques will also add considerable explanatory power to help predict credit risk in a regression context.

IV.2.2 Corporate Social Irresponsibility and Credit Risk

Similar to default risk, predicting credit risk is important for both researchers and practitioners for numerous reasons. Credit risk is the probability of a debtor not repaying the principal and interest on the debt in part or in full (Vassalou and Xing, 2004). Being unable or unwilling to repay all debts is equivalent to a default. Thus, credit risk is also an *ex-ante* evaluation of firm default risk. The first comprehensive theory and model of the pricing of corporate debt (i.e., credit risk) was developed by Merton (1973, 1974), whose work was later advanced by Leland (1994). These models not only enabled pricing any corporate liability but were also used to predict bankruptcy (e.g., Bharath and Shumway, 2008).

Merton's (1974) and Leland's (1994) models have also contributed to continually developing the credit risk analyses of credit rating agencies. For market participants, these rating agencies became an interesting source of credit risk assessments as they provided information

[54] Some of the work in predicting default risk even considers advanced hybrid ML algorithms as a superior method (Yeh et al., 2014). However, we will not discuss these studies due to their specificity and restricted applicability.

about the downside risk of firms no longer being able to repay their debts in part or in full. Besides their importance for market participants, credit ratings are also one of the most important drivers of capital structure choice by executives (Graham and Harvey, 2001). Changes in credit ratings may result in very concrete costs such as changes in coupon rates, losing a contract, or triggering the repurchase of bonds (Kisgen, 2006). Similar to previous research (e.g., Sun and Cui, 2014; Hsu and Chen, 2015; Kealhofer, 2003), our study uses these credit ratings as a proxy for credit risk. Credit risk is economically relevant in various dimensions. Firms with higher distress and credit risk offer significantly lower financial returns (Dichev, 1998; Campbell et al., 2008). Moreover, credit risk has been linked to market illiquidity and increased yield spreads (Ericsson and Renault, 2006). Ultimately, higher credit risk hurts firm profitability due to higher refinancing costs. Therefore, credit risk is key not only to the decision making processes of lenders but also to those of investors.

In addition to the importance of credit risk, the impact of CSP on credit risk has attracted quite some attention. CSP is multidimensional and thus covers firm behavior in the environmental (e.g., pollution control), social (e.g., diversity), and corporate governance (e.g., stakeholder strategy) dimensions (Waddock and Graves, 1997). CSP has been shown to reduce the risk of falling into default (e.g., Sun and Cui, 2014) and firm risk (e.g., Jo and Na, 2012; Lee and Faff, 2009). The aforementioned research suggests that the risk-reducing effect of CSP is also reflected in lower credit risk. In fact, rating agencies tend to reward socially and environmentally responsible firms with higher credit ratings (Hsu and Chen, 2015; Jiraporn et al., 2014; Kiesel and Lücke, 2019). In particular, high performance on ESG aspects concerning primary stakeholders (i.e., employees, customers, etc.) plays a major role in higher credit ratings (Ashbaugh-Skaife et al., 2006; Attig et al., 2013). Anecdotal evidence exists for cases in which credit ratings were alternated due to changes in the CSP of firms such as Wells Fargo or Toshiba in the period 2016-2018 (Henisz and McGlinch, 2019).

Another stream of research has investigated the spreads of bonds and bank loans (henceforth, credit spreads) as a measure of credit risk. Several studies have shown that firms with high (low) CSP are rewarded (penalized) by the corporate bond market with lower spreads (Oikonomou et al., 2014; Goss and Roberts, 2011; Chava, 2014; Stellner et al., 2015; Drago et al., 2019). On the one hand, less responsible firms pay between 7 and 18 basis points higher

bond risk premia (Goss and Roberts, 2011). Moreover, firms that have environmental concerns are charged significantly higher interest rates on bank loans (Chava, 2014). On the other hand, performing well on ESG aspects such as "support for local communities, higher levels of marketed product safety and quality characteristics, and avoidance of controversies regarding the firm's workforce" positively impacts bond risk premia (Oikonomou et al., 2014). Moreover, high CSP is linked to reduced zero-volatility spreads (Stellner et al., 2015). The announcement of ESG performance ratings has also been shown to decrease CDS spreads, pointing to the credit risk-decreasing role of ESG performance (Drago et al., 2019).

However, the evidence on CSP and credit risk is rather mixed. The same study that found reduced zero-volatility spreads due to high CSP did not find a statistically significant effect of high CSP on credit ratings (Stellner et al., 2015). In terms of credit spreads, another study found that *more* responsible firms have to pay *higher* bond risk premia (Menz, 2010). However, due to the lack of statistical significance, Menz (2010) concluded that environmental and social factors have not yet been considered in corporate bond pricing. Moreover, both Stellner et al. (2015) and Menz (2010) focused on the European market, which might explain their findings.

While the relationship between positive CSP (i.e., ESG performance) and credit risk is well explored, a more recent study has shifted the focus from CSP to CSI (i.e., negative ESG performance). CSI can be defined as a "set of corporate actions that negatively affects an identifiable social stakeholder's legitimate claims" (Strike et al., 2006, p. 852). It comprises behavior that is explicitly irresponsible, while the absence of this behavior is not necessarily responsible (Strike et al., 2006). For example, while the violation of human rights is generally perceived as irresponsible, *not* violating human rights, however, should be a matter of course. This observation also explains why corporate socially responsible behavior is hardly reported in the news, whereas socially irresponsible behavior receives high media coverage (Kölbel et al., 2017).

Kölbel et al. (2017) investigated whether CSI is linked to higher financial risk. Contrary to the idea of a value-enhancing and reputation-building effect of CSP (i.e., ESG performance), they instead examined the risk-generating effects of CSI, based on research by Strike et al. (2006). To measure CSI, they used data from RepRisk. The link between CSI and financial risk

is promising as both concern the risk dimension of firms. Kölbel et al. (2017) found that firms facing CSI also experience significantly higher CDS spreads and that this effect is primarily driven by the governance dimension.

Our investigation is based primarily on Kölbel et al. (2017)'s findings on CSI and CDS spreads. Their results lead us to expect that information on CSI should be relevant to predicting credit risk and thus be priced in credit ratings. In contrast to previous studies, we are interested in robust empirical evidence as to whether CSI (i.e., compared to ESG performance) informs credit risk predictions as reflected in credit ratings (i.e., compared to CDS spreads). Similar to Kölbel et al. (2017), we used data from RepRisk to operationalize corporate social irresponsibility. Moreover, we are also interested in the predictive ability that a machine learning regression design can add to this kind of prediction problem.

IV.3 Research Design

This section discusses CSI predictors, additional market- and accounting-based predictors, the measure of credit risk (i.e., the target), and the sample collection process. The data part is followed by explaining the ML algorithms used in this study.

IV.3.1 Predictors for Corporate Social Irresponsibility

We used data from RepRisk to measure the CSI of firms. RepRisk specializes in ESG and business conduct risk research by leveraging artificial intelligence and big data to derive relevant risk metrics. Therefore, RepRisk adheres to the term "ESG risk." RepRisk captures material ESG risks and violations of international standards, which can impact the reputation and financial performance of firms. They screen over 90,000 public sources and stakeholders in 20 different languages on a daily basis. Thus, risk metrics are also updated daily. Their screening universe comprised 130k public and private firms (cross-sector and cross-market) when we conducted our study. The historical data on RepRisk dates back to January 2007. The research scope covers 28 broad, comprehensive, and mutually-exclusive ESG issues. For instance, "waste issues," "local pollution," "forced labor," "human rights," and "fraud." Every risk incident in the

RepRisk database is linked to one of these 28 ESG issues. RepRisk identifies any company that is associated with a negative ESG incident.

RepRisk analyzes each risk incident according to its severity (harshness), reach (influence), and novelty. Finally, using a proprietary algorithm, RepRisk quantifies a firm's ESG risk with its "RepRisk Index (RRI)." The RRI captures and quantifies reputational risk exposure related to the 28 ESG issues. It ranges from 0 to 100, and higher values indicate higher ESG risk exposure. RepRisk provides three different RRI values. First, Current RRI, which measures the current ESG risk exposure (i.e., quarterly in our case). Second, Peak RRI, which captures the highest level of ESG risk exposure over the two previous years. While the Current RRI is instead a snapshot of a firm's exposure to ESG risks, Peak RRI accounts for the time-delayed manifestations of overall ESG risk. Third, RRI Change or Trend, which is the past 30-day increase or decrease of the RRI. Since we used quarterly observations, we only considered Current RRI and Peak RRI as predictors of credit risk. We further used a vector that contains the number of incidents for the 28 single ESG issues multiplied by the respective news count as an alternative measure of CSI.[55]

One of the advantages of RepRisk compared to, for instance, MSCI ESG or Refinitiv is that the RRI is based on third-party disclosed information (i.e., news media) rather than on self-disclosed information. We are, of course, aware that this fact does not necessarily imply an entirely objective measure of CSI. This concern is reinforced by the fact that the algorithm calculating the RRI is proprietary and thus not available for external scrutiny. Moreover, only a few studies have so far used RepRisk data (e.g., Kölbel et al., 2017). We believe this presents a chance but must also be kept in mind as a caveat. Ultimately, we find that RepRisk provides a reliable estimate of CSI (i.e., *not* ESG performance), one that is readily available to researchers and market participants.

[55] More information on our ESG Issues measure can be found in Table B.1. For more information on the RepRisk methodology, please see https://www.reprisk.com/content/static/reprisk-methodology-overview.pdf.

IV.3.2 Additional Accounting- and Market-Related Predictors

In addition to CSI, we controlled for several important factors of credit risk, which are well-founded in the previous literature. We obtained our additional credit risk predictors from Compustat/CRSP for all our sample firms. We follow Agarwal and Taffler's (2008) three main arguments for an accounting-based model: First, credit risk does not suddenly arise but evolves slowly and thus is best reflected in fundamental values. Second, "window dressing" very likely only marginally affects predictions when using a sophisticated set of accounting variables. Third, loan covenants tend to be based on fundamental values, as reflected in accounting-based numbers.

Most importantly, we used the common accounting-based variables for corporate distress as evidenced by Altman (1968) and Ohlson (1980): Working capital/Total assets (LIQU), Retained earnings/Total assets (PROF), Earnings before interest and taxes (Ebit)/Total assets (OPEF), Market value/Long-term debt (ME), and Sales/Total assets (AT). Moreover, we used predictors that account for short-term effects on financial performance and credit risk (Barboza et al., 2017). These variables include growth in assets (ASSETS), growth in sales (SALES), and percentage change in the price-to-book ratio (P/B).

Since firms with higher leverage are also more likely to have difficulties in repaying their debts, we accounted for the predictive ability of long-term debt/common equity (LEV) (Oikonomou et al., 2014; Attig et al., 2013; Sun and Cui, 2014). More mature firms are usually less likely to face default, which is why we included firm age (AGE) as a predictor (Sun and Cui, 2014). Firm age is calculated as the current year minus the year in which the company first appeared in the Compustat/CRSP database. Smaller firms usually experience much higher credit risk than larger firms (Vassalou and Xing, 2004). Therefore, we included firm size (SIZE) as a predictor.

Following Merton (1974) and Ashbaugh-Skaife et al. (2006), we added operating income before depreciation/interest expense (interest coverage; INTCOV) as an early indicator of the ability to repay debt and net property, plant, and equipment (PPE)/total assets (CAPINT).[56]

[56] In contrast to Ashbaugh-Skaife et al. (2006), we used net values for PPE as these were available for both US and

Based on the Global Industry Classification Standard, we also included sector dummies into our models to account for the impact of sector specificity on credit risk. All our variables are calculated as ratios at the end of every quarter, which allowed us to investigate a large international sample without facing currency conversion issues. However, in an additional analysis, we repeated our main analyses with a more homogeneous sample by using only US firms. All predictors were measured at two quarters before the credit rating (i.e., with a lag of two quarters, i.e., at t_{-2}) to measure the *ex-ante* predictive ability of our models.

IV.3.3 Proxy for Credit Risk

Regarding the target variable, our study differs substantially from previous research on default or bankruptcy prediction, which has commonly used actual firm defaults, as shown in Alaka et al. (2018). However, we are interested in firm credit risk that ultimately informs *ex-ante* about firm default probability. We measured credit risk by S&P Global long-term issuer credit ratings (henceforth S&P ratings) as these ratings constitute an overall and forward-looking evaluation of a debtor's creditworthiness (i.e., the capacity and willingness to repay the obligations due).[57] S&P ratings have been widely used as a proxy for firm credit risk (e.g., Sun and Cui, 2014; Hsu and Chen, 2015; Kealhofer, 2003).

In line with Hsu and Chen (2015) and Avramov et al. (2009), we coded the S&P ratings as a numerical variable ranging from 1 to 22. 1 indicates a AAA rating (i.e., the highest possible) and 22 reflects a D rating (i.e., the lowest possible). Thus, higher values for S&P ratings indicate higher credit and default risk, and vice versa. It is important to note that S&P Global claims that their rating process considers corporate sustainability and social responsibility activities.[58] They also claim that ESG aspects are considered along three dimensions in their rating process: business risk (e.g., competitive position), financial risk (e.g., cash flow/leverage assessment), and management and governance. Therefore, when assuming that ESG aspects are reflected in

international firms on Compustat/CRSP.

[57] For detailed information on S&P ratings, please see https://www.standardandpoors.com/en_US/web/guest/article/-/view/sourceId/504352.

[58] For more information on ESG considerations in S&P credit ratings, please see https://www.spglobal.com/ratings/en/products-benefits/products/sustainable-finance and https://www.spglobal.com/ratings/en/products-benefits/products/esg-in-credit-ratings#.

credit ratings, using information about CSI in our models should inform credit risk predictions and thus increase prediction accuracy.

IV.3.4 Sample Development

We began collecting data for CSI from RepRisk by retrieving quarterly observations between March 31st, 2007 (i.e., the inception of the RepRisk database), and September 30th, 2019 (i.e., the latest available data), for all firms in the RepRisk universe. This procedure yielded 805,617 firm-quarter observations with available data. However, the RepRisk universe also contains many private (i.e., unlisted) firms whose accounting- and market-based values are available only to very limited extent. In fact, after obtaining data from Compustat/CRSP on the additional predictors and merging the two datasets, our sample shrank to 502,090 firm-quarter observations. Next, we obtained quarterly S&P Global long-term issuer credit ratings from Compustat - Capital IQ for the same period. If a rating was revised within the same quarter, we used the rating that was closer to the end of the quarter and thus also more up-to-date. After merging credit risk data with RepRisk and Compustat/CRSP data and after dropping observations with missing values, we ended up with an unbalanced final sample of 43,476 firm-quarter observations (representing 1,850 firms).

In a machine learning context, it is vital to further split one's sample into a training sample and a test (or validation) sample. More specifically, our ML models were only trained on the training sample and not on the test sample. Accordingly, we used the test sample to make out-of-sample predictions. For the training sample, we used 80% of our data (i.e., 34,780 firm-quarter observations). The test sample consisted of the remaining 20% (i.e., 8,696 firm-quarter observations). We developed a stratified sampling approach, ensuring that the target variable's labels (1 for AAA to 22 for D) were first shuffled and then approximately evenly distributed among the training and test samples.[59] Without stratifying, we would have risked biasing our results by uneven distribution or by the order of the target variable labels due to a purely random split between training and test sample (i.e., sample bias).

[59] We stratified our sample by applying the *allclose* function of NumPy (https://docs.scipy.org/doc/numpy/reference/generated/numpy.allclose.html). The exact code can be found in our Github repository, to which access can be granted on request.

IV.3.5 Prediction Using a Selection of Machine Learning Algorithms

We used Python 3.7.4 and a set of algorithms implemented by the widespread ML library *sklearn* to train all models, except for one, and to make predictions with new and unseen data (i.e., out-of-sample). We used the following algorithms: Linear regression, ElasticNet, Ridge, OrdinalRidge, Support Vector Regression (SVR) with a linear kernel (LinearSVR), SVR with a radial basis function (RBF) kernel (RbfSVR), Decision Tree regression (DTR), Random Forest regression (RFR), AdaBoost regression (AdaBoostR), Gradient Boosting regression (GBR), K-nearest Neighbors regression (KNNR), and Multi-layer Perceptron regression (MLPR). While the first eleven algorithms belong to the typical ML repertoire, the last algorithm, MLPR, is a deep learning technique. Algorithms that do not explicitly account for non-linearity in the data (i.e., Linear regression and LinearSVR) or simply apply a regularization term (i.e., ElasticNet, Ridge, and OrdinalRidge) are considered "traditional" algorithms in this study. The remaining algorithms are considered some sort of "advanced" ML algorithms. Below, we described the applied algorithms in more detail.

Linear Regression

Our first algorithm is "plain vanilla" linear regression (or ordinary least squares (OLS) regression). By fitting a linear model's coefficients to the training data, we minimized the residual sum of squares between the actual output and the approximated output.

Elastic Net

Elastic Net is a regularization technique that adds a cost term to linear regression. Elastic Nets can be useful when only a few predictors are relevant or when the predictors are highly correlated. Elastic Net simply shrinks the weights of the irrelevant predictors. Since default and credit risk is usually determined by a considerable number of potential factors (e.g., accounting- and market-related data as well as non-financial data), we ran an Elastic Net algorithm on our extensive sample. This approach allowed us to test whether only a subset of predictors already yields reasonable results. Elastic Net is a common regularization technique that has also been applied in finance (e.g., Wu and Yang, 2014).

Ridge

Ridge regression is another regularization technique that applies a penalty to the size of the coefficients, usually increasing their robustness against collinearity and avoiding over-fitting on the training data.

Ordinal Ridge

Similar to Ridge regression, Ordinal Ridge applies the $\ell 2$ penalty for regularization. However, the main difference to Ridge regression is that the output value is ordinal (i.e., a scale of categories where the relative ordering matters). Ordinal regression models explicitly account for data ordinality (McCullagh, 1980). Ordinal Ridge seems to be a promising algorithm for our type of optimization problem since we are measuring credit risk continuously, yet on a predetermined scale from 1 to 22.[60]

Support Vector Regression

Support Vector Regression (SVR) or Support Vector Machines (SVM) are another very common ML algorithm with various possible applications, for instance, bankruptcy prediction (Chen, 2011). *sklearn* implements ε-SVR based on the LIBSVM library from Chang and Lin (2011) and the LIBLINEAR library from Fan et al. (2008).[61] SVR uses kernel functions to solve the optimization problem. We used two SVR algorithms: one with a linear kernel (LinearSVR), and another with a radial basis function (RBF) kernel (RbfSVR). In general, SVR fits a hyperplane on the training data, which maximizes the margin (i.e., as many training instances within the margin, restricted by the margin of tolerance, defined as ε) and thus minimizes the error.

Decision Tree Regression

In contrast to, for instance, linear regression, Decision Tree regression (DTR) is a non-

[60] We used the *mord* package to implement Ordinal Ridge since *sklearn* has no ordinal ridge (or even simple ordinal regression) implementation. For more information, please see https://pythonhosted.org/mord/. More details on the scale of our target variable can be found in Section IV.3.3.

[61] For reasons of consistency, we adhered to the LIBSVM implementation in *sklearn* for RbfSVR **and** LinearSVR. However, it is very likely that using the LIBLINEAR implementation for LinearSVR increases computing efficiency and decreases training time compared to the LIBSVM implementation (Fan et al., 2008).

parametric algorithm. However, it also leverages supervised learning,[62] similar to all the other algorithms, except for the only artificial neural network (ANN) method in our study, Multilayer Perceptron Regression (MLPR). DTR fits a model to the training data that predicts an output value by learning from simple decision rules based on the predictor's data. DTs are easy to understand and interpret. They also require almost no data adjustments (e.g., in contrast to MLPR). *sklearn* uses an optimized version of the Classification And Regression Tree (CART) algorithm, which was first introduced by Breiman et al. (1984).

Random Forest Regression

DTR typically over-fits the training data and thus exhibits high variance when making predictions based on new and unseen data (i.e., the test data). Random Forest regression (RFR) is an ensemble approach. It makes predictions by combining several base estimators (e.g., decision tree regressors) through averaging (also called "bootstrap aggregating" or "bagging"). The combined and averaged prediction obtained by bagging is usually superior to the prediction of a single estimator (i.e., higher accuracy, robustness, and generalizability) (Breiman, 2001). This superiority originates from two sources of randomness in every single estimator: bootstrapped sampling and random subsets of predictors (Breiman, 2001). Bootstrapped sampling ensures that every estimator (e.g., decision tree regressor) is trained on a different draw (with replacement) from the sample.

The increased randomness implicates more decoupled prediction errors between the single estimators; thus, trading decreased variance for slightly increased bias.[63] Ultimately, bagging reduces over-fitting the model on the training data (i.e., higher generalizability).

AdaBoost Regression

In contrast to the bagging algorithm RFR, AdaBoost regression (AdaBoostR) uses a technique called "boosting." In boosting algorithms, the single estimators are built sequentially rather than independently or in parallel (as in bagging algorithms). Therefore, boosting fits a series of so-called "weak learners" (e.g., small decision tree regressors) on repeatedly adjusted

[62] For a detailed explanation of supervised vs. unsupervised learning, please see Basheer and Hajmeer (2000).

[63] For a discussion on the bias-variance trade-off, please see Géron (2017, p. 129).

training data (Hastie et al., 2009). At each iteration, the boosting algorithm learns from the errors made by the previous weak learner and improves the prediction. Thus, samples that are difficult to predict become increasingly influential with every iteration (Hastie et al., 2009). The general idea of AdaBoostR is that the weights of the various instances are adjusted at every iteration (Géron, 2017).

Gradient Boosting Regression

In addition to AdaBoostR, we decided to use Gradient Boosting regression (GBR) as an alternative algorithm using the boosting technique. In contrast to AdaBoostR, at every iteration, GBR fits a new predictor on the residual errors of the previous predictor.

K-nearest Neighbors Regression

K-nearest Neighbors regression (KNNR) is another very intuitive and efficient regression algorithm. It makes predictions based on new data by relying on the similarity of predictors (or features) of data points from the training data. KNNR predicts a value based on how closely it can be resembled with the points in the training data. If k equals 5, KNNR uses the average of the five closest neighboring data points as a final prediction.

Multi-layer Perceptron Regression

The last algorithm is Multi-layer Perceptron regression (MLPR), which is a DL technique. We used the widespread and well-proven algorithm implementation for first-order optimization problems called Adam, developed by Kingma and Ba (2015). Adam uses stochastic optimization only requiring first-order gradients. Adam is very memory-efficient, and parameter updates occur independently of rescaling the gradient (Kingma and Ba, 2015).

Most importantly, and similar to previous research (e.g., Barboza et al., 2017), we used default settings for all algorithms. Training and prediction time did not exceed one hour in total, given our tidy data set and an efficient algorithm implementation with *sklearn*. Similar to Barboza et al. (2017), none of the algorithms needed more than six minutes for training and prediction. More information on the optimization problems and algorithm settings can be found in Table C.13 and C.1.

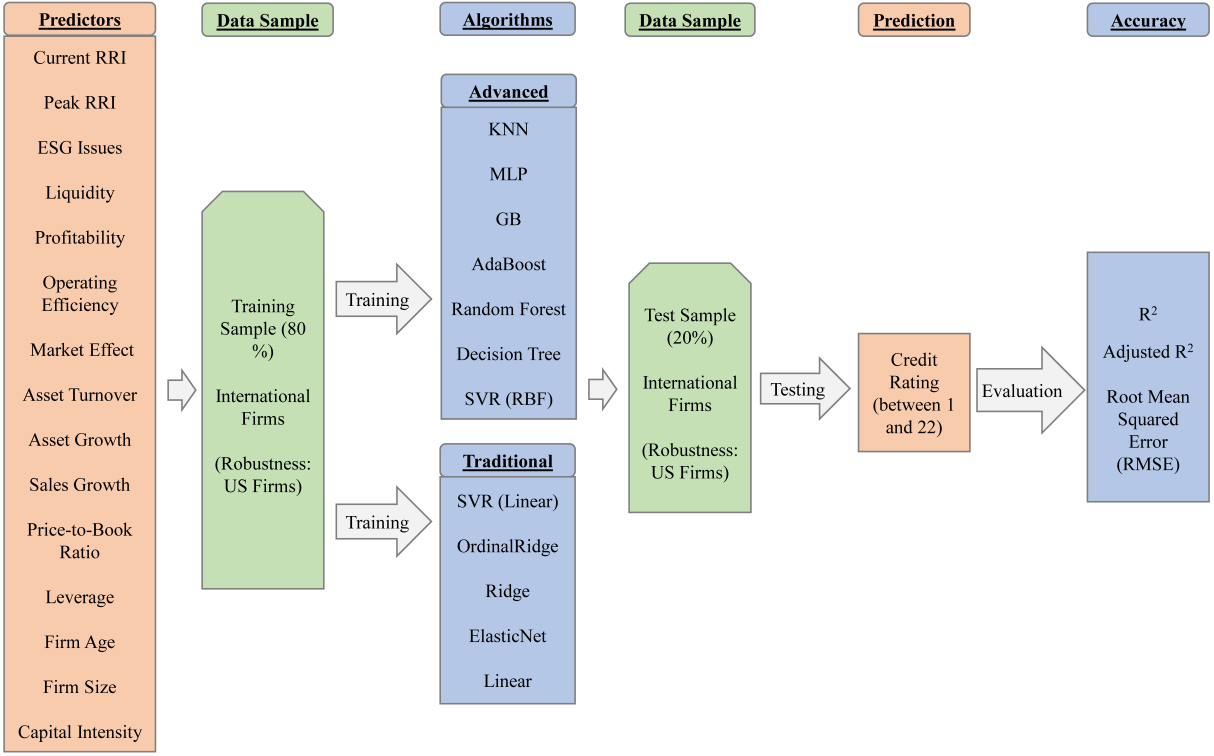
We followed previous research and explicitly refrained from data adjustments such as standardization, normalization, or the removal of outliers (e.g., Barboza et al., 2017; Cleofas-Sánchez et al., 2016). In addition to using default package settings of ML algorithms, this approach not only increases replicability but also enables drawing conclusions about practical applicability. However, it is very likely that the predictive ability of some of our models considerably suffered under this approach. The two exceptions were SVR (linear and RBF kernel) and MLPR, for which we used standardized data (i.e., removed the mean and scaled to unit-variance). Previous research has shown that predictor scaling (i.e., standardization) is a crucial requirement for algorithms such as SVR and MLPR (e.g., Shanker et al., 1996). Unreported results have shown that using non-standardized predictors in SVR and MLPR may lead to highly biased and arbitrary results. Moreover, training time is much higher with non-standardized predictors, which ultimately led us to retain the standardized version as replication needs less computing power. Figure 5 illustrates the process of training and testing in our machine learning setting.

Most importantly, and similar to previous research (e.g., Barboza et al., 2017), we used default settings for all algorithms. Training and prediction time did not exceed one hour in total, given our tidy data set and an efficient algorithm implementation with *sklearn*. Similar to Barboza et al. (2017), none of the algorithms needed more than 6 minutes for training and prediction. More information on the algorithms and settings that we used, as well as the optimization problems, can be found in Table C.13 and C.1 in the appendix.

We followed previous research and explicitly refrained from data adjustments such as standardization, normalization, or the removal of outliers (e.g., Barboza et al., 2017; Cleofas-Sánchez et al., 2016). In addition to using default package settings of ML algorithms, this approach not only increases replicability but also enables drawing conclusions about practical applicability. However, it is very likely that the predictive ability of some of our models considerably suffered under this approach. The two exceptions were SVR (linear and RBF kernel) and MLPR, for which we used standardized data (i.e., removed the mean and scaled to unit-variance). Previous research has shown that predictor scaling (i.e., standardization) is a crucial requirement for algorithms such as SVR and MLPR (e.g., Shanker et al., 1996). Unreported results have shown that using non-standardized predictors in SVR and MLPR may lead to highly

biased and arbitrary results. Moreover, training time is much higher with non-standardized predictors, which ultimately led us to retain the standardized version as replication needs less computing power. Figure 5 illustrates the process of training and testing in our machine learning setting.

Figure 5: Graphical Approach



Note: This figure shows the training and testing process using all relevant predictors and algorithms. The model specifications are not part of this illustration as the same training and testing procedure was used for all specifications.

We solved the credit risk prediction task by using the three aforementioned distinct predictors (Current RRI, Peak RRI, and the vector of ESG issues) and by leveraging all of the explained ML algorithms. In each of the three specifications, we included a vector of the additional accounting- and market-related predictors and a vector of sector dummies. We benchmarked the three specifications against a baseline specification that *only* consisted of the accounting- and market-related predictors and sector dummies. We tested for model accuracy using the three most common measures for regression designs: R^2, adjusted R^2, and RMSE.

Similar to Gu et al. (2020), we also tested the pairwise differences in prediction accuracy for statistical significance according to the Diebold and Mariano (1995) test (henceforth, DM test).[64] We made two types of comparison: First, we compared differences in accuracy between the predictions of algorithms (e.g., Linear regression vs. Random Forest regression) but within our baseline model (i.e., without considering CSI). Second, we compared differences in accuracy between the predictions of our models (e.g., baseline vs. Current RRI) but within a certain algorithm (e.g., Linear regression). In both cases, we defined the test statistic between two prediction samples $DM_{12} = \frac{\bar{d_{12}}}{\sigma_{\bar{d}_{12}}}$, where

$$\bar{d_{12}} = \frac{1}{T-1} \sum_{t=1}^{T} \left[(e^{(1)})^2 - (e^{(2)})^2 \right] \tag{IV.3.1}$$

$(e^{(1)})^2$ and $(e^{(2)})^2$ are the prediction errors of both tested methods and $\sigma_{\bar{d}_{12}}$ denotes the standard deviation.

IV.4 Descriptive Statistics

Table 14 shows the mean, median, standard deviation, minimum, and maximum of the target variable and the predictors used in our models. According to RepRisk, the RRI of most international firms is located somewhere between 26 and 49 (medium risk exposure) due to their global footprint, and only very few firms reach an RRI between 75 and 100 (extremely high-risk exposure). In our sample, the mean and median of Current RRI and Peak RRI are below 25, which

[64] Our DM test includes the modifications suggested by Harvey et al. (1997).

is the threshold for "medium ESG risk exposure" according to RepRisk. However, standard deviation is comparably high (13.9 and 18.0 points, respectively). The minimum and maximum values for Current RRI and Peak RRI are 0 and 79, respectively. Zero represents firm-quarter observations with "no ESG risk exposure" while 79 represents observations with "extremely high ESG risk exposure." Therefore, the descriptives indicate adequate variation in our data and thus the potentially good explanatory and discriminatory power of our two main predictors. The remaining predictor values are comparable with other studies (e.g., Barboza et al., 2017; Oikonomou et al., 2014). However, as mentioned, we refrained from data adjustments (e.g., removing outliers or scaling), as reflected in the summary statistics.

Table 14: Summary Statistics – Predictors and Target

	Mean	SD	Min	50%	Max
Credit Risk	10.681	3.193	1.000	10.000	22.000
Current RRI	13.466	13.856	0.000	13.000	79.000
Peak RRI	24.539	18.048	0.000	27.000	79.000
LIQU	0.100	0.148	−4.756	0.080	0.862
PROF	0.119	0.451	−9.649	0.136	2.184
OPEF	0.022	0.020	−0.360	0.019	0.458
ME	0.844	2.161	−12.189	1.001	16.801
AT	0.208	0.163	−0.127	0.166	2.245
ASSETS	0.233	42.616	−1.000	0.006	8885.494
SALES	0.244	41.644	−5.480	0.010	8682.722
P/B	0.248	59.139	−3384.727	−0.011	8303.042
LEV	6.163	746.361	−7871.500	0.689	109761.545
AGE	27.661	17.999	0.000	22.000	69.000
SIZE	9.289	1.896	3.957	8.952	19.694
INTCOV	1.993	1.113	−6.444	1.961	9.588
CAPINT	0.352	0.256	0.000	0.300	0.989

Notes: This table reports time-series averages of cross-sectional means and standard deviations (SD), the minimum (Min), the 50th percentiles (Median) and the maximum (Max) for all variables. The number of observations (N) is 43476.

Table 15 shows the summary statistics for all 28 ESG issues covered by RepRisk. The three issues "Corruption, bribery, extortion and money laundery," "Fraud," and "Violation of national legislation" had the highest single news count in at least one firm-quarter observation (30, 22, and 28, respectively). A value of 30 for "Corruption, bribery, extortion and money laundery" indicates that this specific firm was involved in or linked to 30 different risk incidents during the quarter. The same applies to the other ESG issues. Of course, these values are

extremes, and the mean and median show that the majority of firms experienced only very few incidents around all ESG issues. However, these extremes show that single firms may be massively exposed to CSI, which in turn may considerably jeopardize firm reputation, firm performance, and solvency. Apart from that, the sample firms' incidents seem to be quite evenly distributed among the remaining ESG issues.

IV.5 Empirical Results

This section presents the main results of our study, the results of our series of additional analyses, and some additional analysis.

IV.5.1 Main Results

Table 16 shows the results for all machine learning algorithms for our baseline specification (i.e., without including CSI exposure). Unsurprisingly, the linear regression algorithm reached only a comparably low out-of-sample prediction accuracy of 56.31% for adj. R^2. Regularization, i.e., ElasticNet, Ridge, and OrdinalRidge, did not increase prediction accuracy after all. Shrinking the weights of supposedly irrelevant predictors, such as in ElasticNet, even impaired prediction accuracy (reducing adj. R^2 from 56.31% to 44.92%). This finding indicates that many of the well-documented predictors used in this study are useful for predicting credit risk and should therefore be considered in credit risk models. While LinearSVR is supposedly unable to capture non-linearity in the data (55.36% for adj. R^2), the comparably poor result for AdaBoostR is surprising (57.07% for adj. R^2). Moreover, our results show that more advanced ML algorithms clearly outperformed linear regression. SVR with a radial basis function kernel and gradient boosting regression were superior to linear regression and the regularization techniques. Out-of-sample adj. R^2 increased to 76.79% for RbfSVR and 73.39% for GBR.

Four of the more advanced ML algorithms achieved adj. R^2s greater than 80%: Decision Tree regression, Random Forest regression, K-nearest Neighbors regression, and Multi-layer Perceptron regression. Random Forest regression achieved the highest out-of-sample prediction accuracy of all algorithms: 89.65% for adj. R^2 and 1.02 for RMSE. This outstanding result

Table 15: Summary Statistics – ESG Issues

	Mean	SD	Min	50%	Max
Animal mistreatment	0.003	0.055	0.0	0.0	2.0
Anti-competitive practices	0.041	0.268	0.0	0.0	7.0
Child labor	0.008	0.106	0.0	0.0	4.0
Climate change, GHG emissions, and global pollution	0.024	0.201	0.0	0.0	8.0
Controversial products and services	0.024	0.240	0.0	0.0	13.0
Corruption, bribery, extortion and money laundery	0.048	0.463	0.0	0.0	30.0
Discrimination in employment	0.009	0.108	0.0	0.0	3.0
Executive compensation issues	0.006	0.085	0.0	0.0	3.0
Forced labor	0.009	0.118	0.0	0.0	7.0
Fraud	0.041	0.332	0.0	0.0	22.0
Freedom of association and collective bargaining	0.011	0.117	0.0	0.0	3.0
Human rights abuses and corporate complicity	0.049	0.312	0.0	0.0	18.0
Impacts on communities	0.084	0.449	0.0	0.0	11.0
Impacts on landscapes, ecosystems and biodiversity	0.076	0.460	0.0	0.0	27.0
Local participation issues	0.014	0.144	0.0	0.0	4.0
Local pollution	0.061	0.404	0.0	0.0	25.0
Misleading communication	0.014	0.140	0.0	0.0	4.0
Occupational health and safety issues	0.032	0.227	0.0	0.0	10.0
Other environmental issues	0.000	0.005	0.0	0.0	1.0
Other issues	0.000	0.011	0.0	0.0	1.0
Other social issues	0.000	0.005	0.0	0.0	1.0
Overuse and wasting of resources	0.005	0.081	0.0	0.0	4.0
Poor employment conditions	0.034	0.240	0.0	0.0	10.0
Products (health and environmental issues)	0.045	0.374	0.0	0.0	17.0
Social discrimination	0.003	0.056	0.0	0.0	2.0
Supply chain issues	0.043	0.322	0.0	0.0	12.0
Tax evasion	0.006	0.091	0.0	0.0	4.0
Tax optimization	0.005	0.095	0.0	0.0	7.0
Violation of international standards	0.007	0.090	0.0	0.0	3.0
Violation of national legislation	0.167	0.741	0.0	0.0	28.0
Waste issues	0.017	0.148	0.0	0.0	5.0

Notes: This table reports time-series averages of cross-sectional means and standard deviations (SD), the minimum, the 50th percentiles (Median) and the maximum for all ESG issues. The number of observations (N) is 43476.

is most likely due to the averaging technique of the Random Forest algorithm. Random Forest regression combines the prediction of several base estimators with an ensemble prediction, which is less likely to over-fit the noise in the training data and thus to generalize better to (unseen) test data. In contrast, Decision Tree regression largely over-fitted the training data (adj. R^2 of 99.90% on the training data compared to 80.72% on the test data). The over-fitting tendency of Decision Tree regression and Random Forest regression has been documented in many previous studies (e.g., Barboza et al., 2017; Piramuthu, 2006). However, Random Forest regression still seems to generalize best to unseen data (i.e., out-of-sample).

K-nearest Neighbors regression and Multi-layer Perceptron regression also seem to be suitable algorithms for predicting credit risk by using the similarities between different firms. After all, they reached an adj. R^2 of 85.05% and 80.35%, respectively. Whereas Barboza et al. (2017) showed that boosting *and* bagging algorithms[65] have a reasonable performance in classification-based bankruptcy prediction, we only found evidence for a high prediction accuracy of bagging algorithms in the case of regression-based credit risk prediction. However, consistent with the findings of Barboza et al. (2017) for Random Forest *classification*, we found that Random Forest also provides the highest prediction accuracy in a *regression* and credit risk design. Other studies have found similar evidence for Random Forests, yet in different contexts than bankruptcy or credit risk prediction (e.g., Yeh et al., 2014).

Using the DM test, we further examined whether the prediction accuracy of one algorithm was significantly superior to that of the others (i.e., pairwise comparisons). Figure 6 displays the results for our baseline specification. Positive values indicate that the column algorithm outperforms the row algorithm. Only statistically significant values ($p < 0.01$) are displayed. Unsurprisingly, the regularization techniques (ElasticNet, Ridge, and OrdinalRidge) did not outperform "plain vanilla" linear regression, as already suggested by the lower adj. R^2. ElasticNet was even outperformed by any other algorithm. However, the more advanced ML algorithms—except for AdaBoostR—yielded predictions that were significantly superior to those of linear regression and the regularization techniques. Again, Random Forest regression turned out to offer the highest prediction accuracy in pairwise comparisons. Therefore, the find-

[65] For a brief description of "boosting" and "bagging," please see the relevant sections for AdaBoostR, Gradient Boosting regression, and Random Forest regression in C.1 in the appendix and Barboza et al. (2017).

Table 16: Prediction Accuracy of Machine Learning Algorithms – Baseline Model

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5601	0.5598	2.1186	Linear	0.5644	0.5631	2.1018
ElasticNet	0.4416	0.4412	2.3868	ElasticNet	0.4508	0.4492	2.3600
Ridge	0.5601	0.5598	2.1186	Ridge	0.5644	0.5631	2.1018
OrdinalRidge	0.5526	0.5522	2.1366	OrdinalRidge	0.5562	0.5549	2.1216
LinearSVR	0.5528	0.5524	2.1361	LinearSVR	0.5548	0.5536	2.1247
RbfSVR	0.7658	0.7657	1.5457	RbfSVR	0.7686	0.7679	1.5320
DTR	0.9990	0.9990	0.1031	DTR	0.8078	0.8072	1.3963
RFR	0.9787	0.9787	0.4659	RFR	0.8968	0.8965	1.0230
AdaBoostR	0.5831	0.5828	2.0624	AdaBoostR	0.5719	0.5707	2.0836
GBR	0.7356	0.7354	1.6425	GBR	0.7347	0.7339	1.6403
KNNR	0.9093	0.9092	0.9621	KNNR	0.8510	0.8505	1.2294
MLPR	0.7923	0.7921	1.4558	MLPR	0.8041	0.8035	1.4095

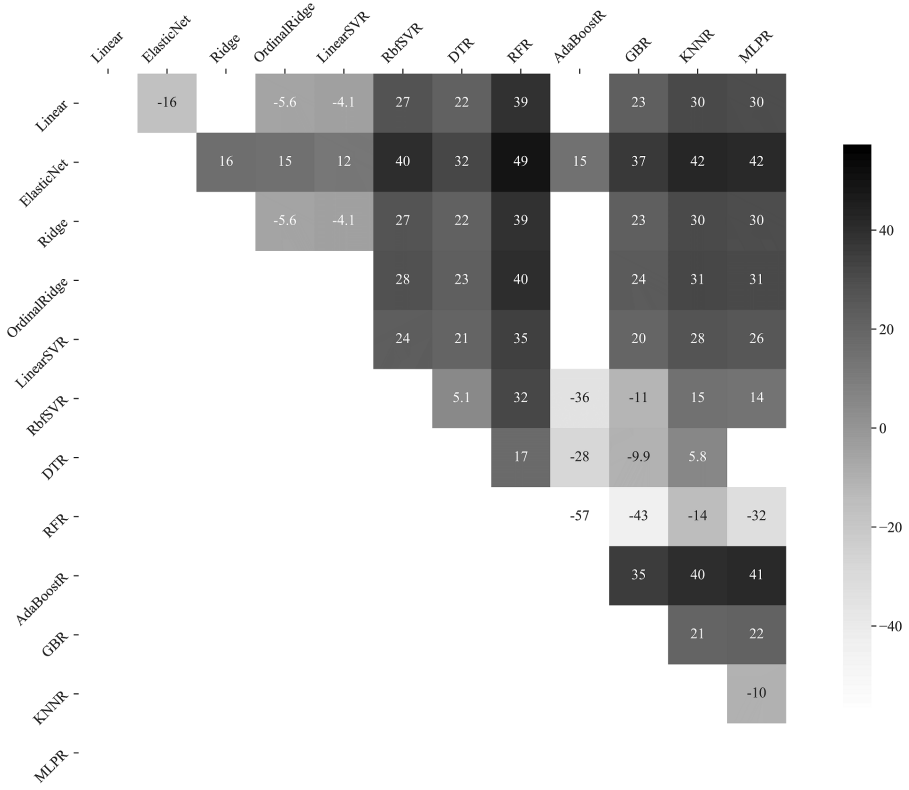
Notes: This table reports the three most common measures of prediction accuracy (R^2, adj. R^2, and RMSE) for the baseline model (i.e., no CSI). The baseline model only consists of the vector of additional accounting- and market-related variables and the vector of sector dummies. For more information on algorithm specifications, please see Appendix C.1 and Table C.13 in the appendix.

ings are in line with our results in Table 16, that advanced algorithms are superior to traditional ones.

Turning to the inclusion of CSI predictors, Panel A of Table 17 displays the prediction accuracy of all algorithms testing the model that includes our first CSI predictor (i.e., Current RRI). The prediction accuracy of some of the algorithms only slightly increased when considering CSI in predicting credit risk. Panel B and Panel C of Table 17 support this finding of slightly increasing prediction accuracy when using Peak RRI and ESG Issues as alternative measures of CSI. Due to the already high prediction accuracy of our baseline specification, most of the explanatory power seems to stem from the accounting- and market-related predictors. Even though previous research has shown that CSI can heavily impact firm performance, firm survival, and even credit risk, we did not find evidence that incorporating CSI into credit risk prediction considerably increases prediction accuracy.

Similar to the pairwise comparison of algorithms for our baseline model specification, we also conducted DM tests for those alternative specifications that include CSI (i.e., Current RRI, Peak RRI, ESG Issues). The two last columns of Table 17 contain the results for the DM

Figure 6: Statistical Significance of Prediction Accuracy – Baseline Model



Note: This figure reports Diebold-Mariano test statistics comparing the out-of-sample predictions between all algorithms for our baseline specification (i.e., no CSI). Positive values indicate outperformance of the column algorithm compared to the row algorithm. Only differences statistically significant at the 1% level or better are displayed. For a description of the algorithms and their abbreviations, please see C.1.

tests, which compare the predictions of the respective alternative specifications to the baseline specification (e.g., Current RRI vs. baseline), holding the algorithm constant. The results were rather mixed. On the one hand, the more traditional ML algorithms (i.e., Linear, ElasticNet, Ridge, and OrdinalRidge) mostly yielded superior predictions when using the alternative specifications compared to the baseline specification (positive t-stat and mostly $p < 0.01$). In other words, predictions improved, yet only marginally. On the other hand, no clear trend could be identified for the more advanced ML algorithms as to whether predictions improved when using the alternative specifications compared to the baseline specifications (e.g., t-stat of -0.36 for RFR using Current RRI and t-stat of 3.85 for RFR using Peak RRI). The most consistent specification turned out to be Peak RRI, which yielded superior predictions compared to the baseline specification in the case of all algorithms—except when using DTR and GBR.

Table 17: Prediction Accuracy – Comparison of Alternative Models

Panel A: Current RRI

Training Sample				Test Sample				DM Test	
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE	t-stat	p-value
Linear	0.5714	0.5711	2.0911	Linear	0.5762	0.5749	2.0731	6.71	0.0000
ElasticNet	0.4501	0.4497	2.3687	ElasticNet	0.4590	0.4574	2.3423	3.20	0.0014
Ridge	0.5714	0.5711	2.0911	Ridge	0.5762	0.5749	2.0731	6.71	0.0000
OrdinalRidge	0.5642	0.5639	2.1086	OrdinalRidge	0.5701	0.5688	2.0880	5.75	0.0000
LinearSVR	0.5652	0.5649	2.1062	LinearSVR	0.5686	0.5673	2.0917	8.08	0.0000
RbfSVR	0.7734	0.7733	1.5204	RbfSVR	0.7758	0.7751	1.5078	6.18	0.0000
DTR	0.9998	0.9998	0.0457	DTR	0.8016	0.8011	1.4183	−1.33	0.1850
RFR	0.9870	0.9870	0.3637	RFR	0.9112	0.9110	0.9489	−0.36	0.7202
AdaBoostR	0.5960	0.5957	2.0301	AdaBoostR	0.5882	0.5869	2.0437	8.95	0.0000
GBR	0.7355	0.7353	1.6426	GBR	0.7351	0.7343	1.6389	0.51	0.6129
KNNR	0.8129	0.8127	1.3817	KNNR	0.7001	0.6992	1.7438	−25.29	0.0000
MLPR	0.8151	0.8150	1.3733	MLPR	0.8182	0.8177	1.3578	7.37	0.0000

Panel B: Peak RRI

Training Sample				Test Sample				DM Test	
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE	t-stat	p-value
Linear	0.5696	0.5693	2.0956	Linear	0.5734	0.5722	2.0799	5.78	0.0000
ElasticNet	0.4487	0.4483	2.3717	ElasticNet	0.4562	0.4545	2.3484	2.22	0.0266
Ridge	0.5696	0.5693	2.0956	Ridge	0.5734	0.5722	2.0799	5.78	0.0000
OrdinalRidge	0.5613	0.5610	2.1156	OrdinalRidge	0.5661	0.5648	2.0977	4.23	0.0000
LinearSVR	0.5632	0.5629	2.1109	LinearSVR	0.5656	0.5643	2.0990	7.06	0.0000
RbfSVR	0.7741	0.7739	1.5182	RbfSVR	0.7770	0.7763	1.5038	7.58	0.0000
DTR	0.9998	0.9998	0.0489	DTR	0.8115	0.8109	1.3826	0.68	0.4983

Table 17 continued from the previous page

RFR	0.9872	0.9872	0.3612	RFR	0.9138	0.9136	0.9349	3.85	0.0001
AdaBoostR	0.5833	0.5830	2.0618	AdaBoostR	0.5768	0.5755	2.0717	2.51	0.0119
GBR	0.7356	0.7354	1.6426	GBR	0.7357	0.7349	1.6372	1.06	0.2889
KNNR	0.8816	0.8815	1.0990	KNNR	0.8086	0.8080	1.3932	−7.93	0.0000
MLPR	0.8142	0.8141	1.3939	MLPR	0.8171	0.8166	1.3618	6.12	0.0000

Panel C: ESG Issues

Training Sample				Test Sample				DM Test	
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE	t-stat	p-value
Linear	0.5681	0.5674	2.0991	Linear	0.5730	0.5702	2.0811	6.99	0.0000
ElasticNet	0.4416	0.4407	2.3868	ElasticNet	0.4508	0.4473	2.3600	NaN	NaN
Ridge	0.5681	0.5674	2.0991	Ridge	0.5729	0.5702	2.0811	6.99	0.0000
OrdinalRidge	0.5610	0.5603	2.1164	OrdinalRidge	0.5665	0.5637	2.0966	5.83	0.0000
LinearSVR	0.5608	0.5601	2.1168	LinearSVR	0.5634	0.5605	2.1043	7.53	0.0000
RbfSVR	0.7157	0.7152	1.7032	RbfSVR	0.7020	0.7001	1.7385	−18.39	0.0000
DTR	0.9990	0.9990	0.0988	DTR	0.8018	0.8005	1.4179	−1.18	0.2362
RFR	0.9865	0.9865	0.3715	RFR	0.9105	0.9099	0.9528	−2.21	0.0269
AdaBoostR	0.5831	0.5824	2.0624	AdaBoostR	0.5719	0.5691	2.0836	NaN	NaN
GBR	0.7364	0.7359	1.6401	GBR	0.7356	0.7339	1.6376	1.30	0.1945
KNNR	0.9000	0.8998	1.0103	KNNR	0.8365	0.8354	1.2877	−7.14	0.0000
MLPR	0.8146	0.8143	1.3753	MLPR	0.7834	0.7820	1.4821	−5.98	0.0000

Notes: This table reports the three most common measures of prediction accuracy (R^2, adj. R^2, and RMSE) for the models, including the CSI predictors. In Panel A, we used Current RRI as a predictor to capture CSI exposure. In Panel B, we used Peak RRI as a predictor to capture CSI exposure. In Panel C, we used the vector of ESG issues as a predictor to capture CSI exposure. The vector of additional accounting- and market-related variables and the vector of sector dummies are included in all models. The statistics from the Diebold-Mariano test are displayed in the last two columns. They test the statistical significance of differences in the accuracy of predictions compared to the baseline specification (i.e., the model does not capture CSI; see Table 16 and Figure 6). Positive t-stat values indicate outperformance compared to the baseline model. "NaN" values indicate that the difference between the predictions was zero, i.e., the models yielded the same predictions. This finding is not surprising since regularization techniques, such as ElasticNet, shrink the weights from irrelevant predictors, which can render an equivalent or highly similar model to the baseline specification. For more information on algorithm specifications, please see Appendix C.1 and Table C.13 in the appendix.

One possible explanation for our results is that our measure of credit risk (i.e., S&P Global long-term issuer credit ratings) does not yet reflect CSI exposure to a larger extent. Another possible explanation is that higher CSI increases the cost of capital, as suggested by Kölbel et al. (2017), instead of being reflected in credit ratings. Therefore, our result does not indicate that CSI is consistently and systematically considered in the credit risk analyses of rating agencies. For example, S&P Global claims that its credit risk analyses consider ESG factors along three dimensions: business risk (e.g., competitive position), financial risk (e.g., cash flow/leverage assessment), and management and governance. However, we cannot rule out that differently operationalizing ESG aspects between RepRisk and S&P Global leads to our results.[66] For example, this difference in operationalization might be due to a possibly higher weighting of the corporate governance dimension, whereas the environmental and social dimensions are underweighted.

Figure 7 illustrates and compares the prediction accuracy of all our algorithms in the training and test samples.[67]

IV.5.2 Additional Analyses – Regression Design

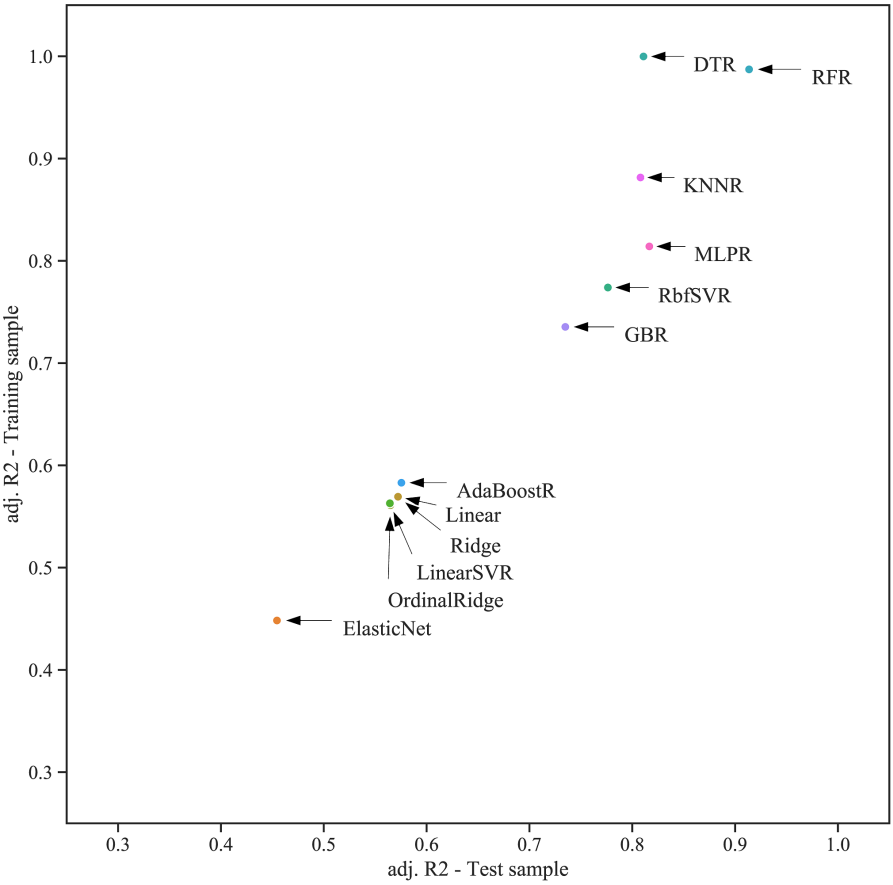
We performed a series of additional analyses in the following. The appendix contains further robustness checks related to an increased test size, an alternative specification, and non-stratified samples.

First, we used only US data instead of global data for our analysis. Table C.14 in the appendix shows that using US data increases the overall prediction accuracy of most ML algorithms. Using Peak RRI, Random Forest regression again yields the highest out-of-sample prediction accuracy with 92.53% for adjusted R^2 and 0.8682 for RMSE. The higher overall prediction accuracy is presumably owed to the lower noise and the higher homogeneity in US data compared to global data.

[66] For S&P Global's claims about integrating ESG in credit risk analysis, please see: https://www.spglobal.com/ratings/en/products-benefits/products/sustainable-finance and https://www.spglobal.com/ratings/en/products-benefits/products/esg-in-credit-ratings#.

[67] We decided only to illustrate the model using Peak RRI rather than the alternative CSI measures, as the model using Peak RRI achieved the highest out-of-sample prediction accuracy overall.

Figure 7: Model Prediction Accuracy – Peak RRI



Notes: This figure shows the prediction accuracy (adj. R^2) of all ML algorithms using Peak RRI as a predictor. The vector of additional accounting- and market-related variables and the vector of sector dummies are included. The y-axis displays adj. R^2 for the training sample and the x-axis shows adj. R^2 for the test sample. For a description of the algorithms and their abbreviations, please see Appendix C.1.

Second, we tried to reduce the noise in the CSI data caused by negligible negative ESG incidents with no or only few (economic) consequences for the firm. More specifically, we followed Kölbel et al. (2017) and controlled for the reach of the news source in which the particular negative ESG incident was reported and for the severity (or harshness) of the negative ESG incident. A higher reach of the news source (e.g., The Financial Times vs. the Nassau Herald) and stronger severity (e.g., large corruption scandal spanning several geographies vs. minor local pollution incident) are more likely to affect credit risk and thus be reflected in credit ratings. Therefore, we calculated interaction terms for our vector of ESG issues. Only negative ESG incidents that appeared in high-reach news sources and with high severity were considered in our ESG Issues predictor. All other negative ESG incidents were weighted zero. As Table C.15 in the appendix shows, this alternative specification of ESG Issues did not considerably change our results. We conclude that negative ESG incidents, which appeared in influential news sources and which had a severely negative impact, do not increase prediction accuracy either and thus are not systematically reflected in credit risk ratings.

Third, we used a time lag of four quarters instead of two quarters for our predictor variables to account for the possibility that it takes longer for CSI to materialize in credit ratings. However, a longer time lag did not qualitatively change our results (cf., Table C.19 in the appendix).

Fourth, we applied a subsample analysis only using firm-quarter observations from the year 2015 and later, based on the observation that credit rating agencies have strengthened their ESG considerations in recent years. Therefore, it is more likely that CSI informs credit risk prediction as reflected in credit ratings during this subperiod. Similar to the results for the full sample, we found that the prediction accuracy of some algorithms slightly improves (e.g., DTR, AdaBoostR, LinearSVR, OrdinalRidge, Ridge, ElasticNet, Linear). In contrast, that of others even slightly declines (e.g., KNNR) due to including CSI. However, in absolute terms, the difference in prediction accuracy between models using CSI predictors and those which do not is only marginal. The effect on prediction accuracy is slightly stronger when only using data since 2015 compared to the full sample (cf., Table C.20 in the appendix). This finding is in line with the recent efforts of credit rating agencies to incorporate CSI more strongly into their credit risk analyses.

Finally, we retrieved both Moody's and Fitch credit risk data from Thomson Reuters to rule out that our result is driven by the use of S&P Global data. Similar to S&P ratings, Moody's and Fitch ratings were converted to numerical values according to the respective scale. Using Moody's and Fitch rating data considerably reduced our sample to 8,202 and 4,261 firm-quarter observations, respectively. Again, the two samples were split into a training (80%) and a test (20%) sample.

The findings indicate that the prediction accuracy of traditional algorithms considerably declines, whereas the more advanced algorithms largely maintain their prediction accuracy scores (cf., Table C.21 and Table C.22 in the appendix). Adjusted R^2s decrease considerably to only slightly above 20% for most algorithms (except when using ElasticNet, which is even well-below 20%). While most more advanced ML algorithms maintain their comparably high out-of-sample prediction accuracy, only Random Forest regression, K-nearest Neighbor regression, and Multi-layer Perceptron regression achieve outstanding out-of-sample prediction accuracy close to or above 80% for adj. R^2. This finding is somewhat surprising since algorithms like Random Forest regression usually work best with sample sizes well-beyond 10,000. In our case, and despite the considerably decreased sample size, the mentioned algorithms maintain their prediction accuracy scores. Overall, S&P ratings are best explained by our models, followed by Fitch ratings, followed by Moody's ratings.

IV.5.3 Predicting Investment Grade vs. Non-Investment Grade – Classification Design

So far, we have used a regression design to predict a credit risk continuum between 0 and 22, which possibly lacks discriminatory power in terms of distinguishing observations with different CSI. To alleviate this concern, we used an alternative classification design to predict whether a firm will be rated investment grade or non-investment grade in the future. In contrast to a regression design, classification predicts whether a particular status applies or not (i.e., a categorical decision between 1 and 0). The most common classification algorithm is Logistic regression.

From both an investor's and an analyst's perspective, it would be very interesting to investigate the prediction accuracy of ML algorithms for classifying a firm into investment

grade or non-investment grade. The non-investment grade status of an asset is an essential piece of information for investors and analysts alike. For example, credit rating decisions such as a non-investment grade status have signaling value for investors and may significantly impact the capital structure of firms (Graham and Harvey, 2001; Kisgen, 2006). Moreover, while CSI did not inform the prediction of credit risk as reflected in credit ratings in a regression design, it potentially informs a credit risk agency's decision to classify a firm as non-investment grade. Therefore, we conducted an additional analysis to test whether our models can reliably predict the non-investment grade status of a firm two quarters into the future and whether data on CSI informs these predictions.

Usually, a firm with a rating of BBB– or lower has non-investment grade status. Accordingly, we classified each firm-quarter observation with a rating of BBB– or lower as non-investment grade (i.e., a value of 1). All other firm-quarter observations with ratings above BBB– were classified as investment grade (i.e., a value of 0). As for our traditional algorithms, we used Logistic regression (Logit) and Support Vector Classification (SVC) with a linear kernel (LinearSVC) as our benchmark classification algorithms. As for our more advanced algorithms, we used SVC with an RBF kernel (RbfSVC), Decision Tree classification (DTC), Random Forest classification (RFC), AdaBoost classification (AdaBoostC), Gradient Boosting classification (GBC), K-nearest Neighbors classification (KNNC), and Multi-layer Perceptron classification (MLPC).[68] Again, we refrained from data adjustments such as standardization, normalization, or the removal of outliers, except for SVC and MLPC. We used four common measures for prediction accuracy in a machine learning classification context (Precision, Recall, F1-Score, and Accuracy), which are explained in detail in Table 18.

The results of our additional analysis are displayed in Table 18. In general, the results of the classification design are in line with those of the regression design. The two traditional ML algorithms, Logit and LinearSVC, exhibit reasonable out-of-sample prediction accuracy of 75.70% and 81.85%, respectively. However, the more advanced ML algorithms exhibit superior prediction accuracy. For example, DTC and RFC have a prediction accuracy of 92.54% and 95.04%, respectively. This result is in line with results for default prediction using a ma-

[68] Due to the similarity of most classification algorithms with the regression algorithms, we refrain from further explanation. However, detailed explanations and blueprints of the algorithms can be found at https://scikit-learn.org/stable/supervised_learning.html.

chine learning classification design (Barboza et al., 2017). Overall, our algorithms show higher Precision, i.e., the accuracy of positive predictions, than Recall, i.e., the fraction of positive observations that are correctly detected by the classifier. The deep learning method MLPC has the highest Recall of 94.37%, which indicates that this method reliably identifies almost 95% of all non-investment grade firms (out-of-sample).

Moreover, we do not find that including CSI consistently increases prediction accuracy when classifying firms into investment grade or non-investment grade. The prediction accuracy of most algorithms either remains approximately the same (e.g., DTC or GBC) or even slightly declines (e.g., Logit or KNNC). Overall, Random Forest classification again dominates the other algorithms with prediction accuracy between 93.99% and 95.06%. These results are in line with our main analysis. Random Forest algorithms provide the highest prediction accuracy for credit risk as measured by credit ratings *and* for classifying firms into non-investment grade. Moreover, including CSI in models does not considerably increase either the accuracy of credit risk prediction or the accuracy of classifying firms into non-investment grade.

Table 18: Additional Analysis – Investment Grade vs. Non-Investment Grade

Panel A: Baseline – Without CSI Exposure

Training Sample					Test Sample				
Algorithm	Precision	Recall	F1-Score	Acc.	Algorithm	Precision	Recall	F1-Score	Acc.
Logit	0.7931	0.6782	0.7312	0.7509	Logit	0.7854	0.6848	0.7317	0.7570
LinearSVC	0.8244	0.8003	0.8122	0.8152	LinearSVC	0.8188	0.8025	0.8106	0.8185
RbfSVC	0.9007	0.8740	0.8872	0.8890	RbfSVC	0.8998	0.8728	0.8861	0.8914
DTC	1.0000	0.9981	0.9990	0.9991	DTC	0.9240	0.9216	0.9228	0.9254
RFC	0.9982	0.9952	0.9967	0.9967	RFC	0.9634	0.9330	0.9480	0.9504
AdaBoostC	0.8481	0.8238	0.8358	0.8383	AdaBoostC	0.8406	0.8210	0.8307	0.8381
GBC	0.8869	0.8625	0.8745	0.8764	GBC	0.8741	0.8550	0.8645	0.8703
KNNC	0.9655	0.9596	0.9626	0.9627	KNNC	0.9281	0.9296	0.9289	0.9311
MLPC	0.9710	0.9596	0.9653	0.9655	MLPC	0.9468	0.9437	0.9452	0.9471

Panel B: Current RRI

Training Sample					Test Sample				
Algorithm	Precision	Recall	F1-Score	Acc.	Algorithm	Precision	Recall	F1-Score	Acc.
Logit	0.7859	0.6686	0.7225	0.7435	Logit	0.7768	0.6677	0.7181	0.7464
LinearSVC	0.8221	0.8022	0.8120	0.8145	LinearSVC	0.8206	0.8046	0.8125	0.8204
RbfSVC	0.9026	0.8772	0.8897	0.8914	RbfSVC	0.8996	0.8750	0.8871	0.8922
DTC	1.0000	0.9993	0.9996	0.9996	DTC	0.9196	0.9242	0.9219	0.9242
RFC	0.9988	0.9953	0.9971	0.9971	RFC	0.9611	0.9277	0.9441	0.9469
AdaBoostC	0.8452	0.8241	0.8345	0.8367	AdaBoostC	0.8406	0.8251	0.8328	0.8397
GBC	0.8860	0.8601	0.8728	0.8748	GBC	0.8733	0.8519	0.8625	0.8686
KNNC	0.9334	0.8898	0.9111	0.9133	KNNC	0.8799	0.8338	0.8562	0.8645
MLPC	0.9731	0.9638	0.9684	0.9686	MLPC	0.9528	0.9460	0.9494	0.9512

Table 18 continued from the previous page

Panel C: Peak RRI

Training Sample					Test Sample				
Algorithm	Precision	Recall	F1-Score	Acc.	Algorithm	Precision	Recall	F1-Score	Acc.
Logit	0.7733	0.6067	0.6799	0.7147	Logit	0.7611	0.6133	0.6792	0.7198
LinearSVC	0.8235	0.8009	0.8120	0.8148	LinearSVC	0.8201	0.8053	0.8127	0.8204
RbfSVC	0.9068	0.8785	0.8924	0.8942	RbfSVC	0.9053	0.8769	0.8908	0.8960
DTC	1.0000	0.9992	0.9996	0.9996	DTC	0.9173	0.9230	0.9201	0.9225
RFC	0.9987	0.9967	0.9977	0.9977	RFC	0.9655	0.9311	0.9480	0.9506
AdaBoostC	0.8462	0.8279	0.8369	0.8389	AdaBoostC	0.8381	0.8281	0.8331	0.8395
GBC	0.8857	0.8607	0.8731	0.8750	GBC	0.8725	0.8526	0.8625	0.8684
KNNC	0.9532	0.9440	0.9486	0.9489	KNNC	0.9179	0.9085	0.9132	0.9164
MLPC	0.9726	0.9591	0.9658	0.9661	MLPC	0.9499	0.9415	0.9457	0.9477

Panel D: ESG Issues

Training Sample					Test Sample				
Algorithm	Precision	Recall	F1-Score	Acc.	Algorithm	Precision	Recall	F1-Score	Acc.
Logit	0.7768	0.6831	0.7269	0.7437	Logit	0.7709	0.6896	0.7280	0.7507
LinearSVC	0.8229	0.8028	0.8127	0.8152	LinearSVC	0.8192	0.8032	0.8111	0.8190
RbfSVC	0.8817	0.8616	0.8715	0.8731	RbfSVC	0.8705	0.8436	0.8568	0.8636
DTC	1.0000	0.9983	0.9992	0.9992	DTC	0.9211	0.9180	0.9195	0.9223
RFC	0.9984	0.9955	0.9969	0.9970	RFC	0.9594	0.9144	0.9364	0.9399
AdaBoostC	0.8492	0.8240	0.8364	0.8390	AdaBoostC	0.8385	0.8222	0.8303	0.8374
GBC	0.8873	0.8633	0.8752	0.8770	GBC	0.8753	0.8529	0.8640	0.8701
KNNC	0.9608	0.9556	0.9582	0.9583	KNNC	0.9198	0.9242	0.9220	0.9243
MLPC	0.9721	0.9763	0.9742	0.9742	MLPC	0.9264	0.9342	0.9303	0.9323

Notes: This table reports four common measures of prediction accuracy for all classification algorithms: Precision, Recall, F1-Score, and Accuracy (Acc.). Precision is the accuracy of the positive predictions, i.e., the number of true positives (TPs) divided by the number of TPs plus the number of false positives (FPs). In this context, "the number of ..." always refers to "the number of observations classified as ... by the algorithm." For brevity, we use the shorter form. Recall is the ratio of positive observations that are correctly detected, i.e., the number of TPs divided by the number of TPs plus the number of false negatives (FNs). The F1-Score is the harmonic mean of precision and recall, i.e., the number of TPs divided by the number of TPs plus the number of FNs and FPs divided by two. Accuracy measures the fraction of correct predictions. In Panel A, we used the baseline specification and thus no CSI predictor. In Panel B, we used Current RRI as a predictor to capture CSI exposure. In Panel C, we used Peak RRI as a predictor to capture CSI exposure. In Panel D, we used the vector of ESG issues as a predictor to capture CSI exposure. The vector of additional accounting- and market-related variables and the vector of sector dummies are included in all models.

IV.6 Contributions to Research and Practical Implications

Our research contributes to the literature and has practical implications in at least two ways. First, whereas previous research has relied on statistical inference, we used a comprehensive machine learning design to train models for out-of-sample prediction of credit risk (as measured

by credit ratings). Previous studies have primarily investigated the impact of ESG performance (e.g., Oikonomou et al., 2014; Stellner et al., 2015; Hsu and Chen, 2015) and of CSI (e.g., Kölbel et al., 2017) on credit risk. However, they have neglected the overall prediction of credit risk using data on CSI and the potential of ML algorithms. The only exception is Dorfleitner et al. (2020), from whom our study differs by using data on CSI (i.e., *not* CSP) and a comprehensive series of machine learning algorithms. Predictive models that use advanced machine learning algorithms potentially add value by capturing the inherent non-linearity and complex interaction effects in the data. Credit risk is potentially influenced by a vast number of factors, financial and non-financial, which need to be considered in modeling. For investors and analysts, it is vital to incorporate these factors into credit risk analysis. ML techniques can help obtain the most accurate credit risk prediction, especially if large amounts of data with complex relationships and interaction effects are available. The resource intensiveness of credit risk evaluation and the importance of firm solvency for financial institutions support the data-driven machine learning approach.

Second, previous research has shown that ESG performance is relevant to credit risk (e.g., Oikonomou et al., 2014; Stellner et al., 2015; Drago et al., 2019). At the same time, major rating agencies claim that they account for ESG aspects to some extent in their ratings.[69] However, we found that including data on CSI does not systematically increase the prediction accuracy of credit risk as measured by credit ratings. Therefore, our finding that CSI does not inform credit risk prediction is somewhat surprising. It also contrasts slightly with previous studies, according to which more environmentally and socially responsible firms are rewarded with higher credit ratings (Jiraporn et al., 2014; Attig et al., 2013; Kiesel and Lücke, 2019). We do not find support for the anecdotal evidence that changes in the ESG performance of certain firms during the period 2016-2018 led to changes in credit risk ratings on a systematic basis (Henisz and McGlinch, 2019). Our result is more in line with research on the European market, which has found that higher ESG performance is not significantly related to higher credit ratings (Stellner et al., 2015). However, our finding holds for both a global and a US sample and is related to CSI as compared to CSP (i.e., ESG performance).

A possible explanation for the diverging results is that rating agencies tend to account

[69] For a detailed discussion of this inclusion, please see above.

for positive ESG performance—at least outside of Europe—but not for the downside risk of negative ESG incidents (i.e., CSI). Another possible explanation is that the operationalization of ESG aspects by credit rating agencies significantly differs from our operationalization (i.e., CSI as measured by data from RepRisk). Our finding is also somewhat surprising in the light of results by Kölbel et al. (2017), who found that CSI impacts credit risk through increased CDS spreads. In contrast, our results indicate that higher credit risk due to CSI is not reflected in credit ratings. It seems that while the market already accounts for CSI, CSI does not (yet) play a major role in credit ratings. Therefore, a potential implication for market participants is that common credit ratings might not entirely reflect the environmental and societal business conduct risks (e.g., weaknesses in corporate governance, low product safety, or compromised employee well-being) that firms are increasingly facing. These business conduct risks potentially entail substantial future expenses (e.g., compensation payments) and reputational damages.

Ultimately, then, Peak RRI adds most to prediction accuracy across algorithms. In absolute terms, however, the additional explanatory power is small compared to specifications only using accounting- and market-related predictors. However, this finding likely depends on our measure of credit risk, i.e., credit rating data from S&P Global, Moody's, and Fitch. In light of our results, and considering the potential importance of ESG aspects for credit risk, it seems reasonable to more strongly incorporate CSI into credit risk analysis. Major credit rating agencies such as Fitch have recently stated that they are intensifying their efforts to integrate ESG into credit risk analysis and to be more transparent about integration.[70]

IV.7 Concluding Remarks

We have investigated whether using advanced ML algorithms and data on CSI improves firm credit risk prediction. We therefore also tested the extent to which increased credit risk due to higher CSI is reflected in credit ratings. We found that the use of advanced ML algorithms considerably improves prediction accuracy for credit risk as measured by credit ratings. However, we found no evidence that CSI is systematically reflected in credit ratings. A long series of additional analyses and robustness checks corroborated our results. In an alternative classification

[70] For the respective announcement by Fitch, please see https://www.fitchratings.com/site/pr/10058528.

design, we investigated whether advanced ML algorithms and data on CSI inform the prediction of a firm's non-investment grade status (i.e., a credit rating of BBB– or lower). Whereas advanced ML algorithms again considerably increase prediction accuracy using this classification design, data on CSI are not consistently informative as to whether a firm is predicted to have non-investment grade status.

The results of our study are limited to some extent. First, the time-series availability of CSI and credit rating data is limited. As CSI data availability increases, training ML models will most likely yield more robust systems These systems include ones that are more robust when considering time trends.

Second, studies like ours face the trade-off between using standardized and non-standardized predictors. On the one hand, regularizing and preprocessing data potentially impedes replicating and interpreting the results. On the other, refraining from adjustments such as standardization may yield biased and arbitrary results for certain algorithms. For instance, SVR and MLPR algorithms are highly sensitive to unscaled data. Therefore, we decided to apply standardization where using non-standardized values would have disproportionately biased our results (i.e., for SVR and MLPR) but refrained from standardization where possible (i.e., all the other algorithms).

Third, even though our trained models are ready for use, the actual prediction output and prediction accuracy seem to vary to some extent among different sources of credit risk data (i.e., S&P Global, Moody's, or Fitch). In machine learning, this limitation is well-known as the model dependency on the structure of learning data and further adjustments to the data or algorithms. While we refrained from specific data adjustments and used default algorithm settings, prediction accuracy will most likely vary when using alternative measures for credit risk.

The last limitation leads us to some avenues for future research. While CSI does not (yet) seem to be systematically considered in credit risk evaluations by rating agencies, Kölbel et al.'s (2017) findings indicated that the impact of CSI on credit risk is already priced in by market participants. This suggestion is supported by market-based studies, which have found a bond yield spreads-increasing effect of poor ESG performance (e.g., Goss and Roberts, 2011; Chava,

2014). Therefore, future research could investigate whether ML-based credit risk models can be significantly improved when using an alternative, market-based measure of credit risk (e.g., CDS spreads). While we refrained from algorithm hyper-parameter tuning, future research could also test whether prediction accuracy might be further increased when using k-fold cross-validation and hyper-parameter tuning.

V Overall Concluding Remarks

This chapter concludes the thesis with some final comments on the main results and contributions (Chapter V.1), the general limitations (Chapter V.2), and some avenues of future research (Chapter V.3).

V.1 Main Results and Contribution to the Seminal Literature

The findings and contributions of this thesis have been numerous and sometimes pointing into different directions. Therefore, this section aims at providing an overview of the main findings and contributions and how they related to each other. Investigating the link between CSP and EQ, I find that the materiality of CSP does not significantly relate to the earnings quality of firms. For the accruals quality measure, preliminary findings show that higher material CSP is linked to higher earnings quality. However, this finding is alleviated in robustness checks. In particular, in regulated industries, results show that both material *and* non-material CSP are linked to *higher* earnings quality. Adding to the previous literature (e.g., Kim et al., 2014; Hong and Andersen, 2011), the results suggest that the positive relationship between CSP and EQ is driven by firms from regulated industries. Besides, for the real earnings management measure, results indicate that both material and non-material CSP are linked to *lower* earnings quality. Therefore, the results are mixed when considering the different measures for earnings quality.

This thesis provides insights into the materiality of CSP, the effects of CSP on EQ, and how this materiality can be operationalized through a recalculation of Thomson Reuters ESG scores. The results support previous findings for a positive relationship between CSP and EQ

when using the AQ measure, but not when using the REM measure (e.g., Kim et al., 2014; Hong and Andersen, 2011). Regarding the REM measure, the results are in line with Prior et al. (2008) and Martínez-Ferrero et al. (2016) finding that firms use their CSP to disguise from earnings management. After all, adding the materiality dimension to CSP does not consistently explain EQ. Nonetheless, this thesis shows how the topical issue of materiality can be integrated into Thomson Reuters ESG scores using guidance from SASB. Contributing to the previous literature, the results showed that industry regulation plays an important role in earnings quality. Moreover, and to the best of my knowledge, no previous study has investigated ESG data from Thomson Reuters and such an extensive European sample comprising 542 firms from 27 different industries.

This thesis further shows that positive and negative CSP are informative to the litigation risk of firms and the market impact resulting from litigation. It finds that firms with lower negative CSP show a lower probability of facing class action lawsuits (i.e., litigation risk). The estimated probability of encountering a class action lawsuit is reduced from 3.1% to 2.4% when improving the measure of negative CSP, the ESG controversies score, by one standard deviation. Further, the results show that losses in the market value of a firm with high positive CSP, i.e., high ESG performance, are only half as high as the losses from firms with low or even mediocre positive CSP. In other words, firms with low or mediocre positive CSP face an abnormal loss of US $1.14bn in market value during litigation. A series of robustness checks corroborates this finding. Finally, the implementation of these findings in a trading strategy shows that investors can benefit directly from *monthly* alphas of up to 1.31%.

The contributions to the seminal literature on risk management and stakeholder theory are various (e.g., Harjoto and Laksmana, 2018; Bouslah et al., 2013; El Ghoul et al., 2011). The thesis quantifies how a firm's ESG controversies inform directly about the risk of litigation. Therefore, it shows that running a business responsibly can reduce the number of lawsuits and thus prevent reputational damage and higher cost of capital. This finding implies that socially and environmentally responsible firms are also of interest to investors as the tail event risk of litigation is lower. Moreover, the result of a moderating effect of positive CSP on firm value during class action lawsuits is especially relevant to investors due to shareholder value protection. Investors could consider information about positive and negative CSP in their investment

decisions and thus decrease their exposure towards idiosyncratic risk or shield against losses in market value. More specifically, the suggested trading strategy has delivered a statistically significant above-market return for the years 2014 to 2017—though, within the limits of a backtest. The thesis adds further evidence beyond some studies related to litigation (Hong et al., 2019; Koh et al., 2014; Aouadi and Marsat, 2018) on four dimensions: the prediction of litigation events as a function of ESG controversies, the use of actual litigation events instead of rather crude measures like industry proxies, the derivation of the real market impact on firm value around litigation, and the implementation of the findings in a trading strategy.

The thesis further finds that information on negative CSP (as measured by ESG risks) is not (yet) incorporated systematically and consistently in credit ratings. While advanced ML algorithms improve prediction accuracy for credit risk, including ESG risks into predictive models does *not* consistently increase prediction accuracy. This finding is limited to the extent that it only holds for the historical data (2007 to 2019), and the ESG risks measures from RepRisk used in the study. Major credit rating agencies may use a different operationalization of negative CSP. However, the thesis provides evidence that advanced ML algorithms substantially improve prediction accuracy due to accounting for non-linearity and complex interaction effects in the data. Therefore, investors and analysts can benefit from a synthesis of credit risk prediction with state-of-the-art ML algorithms.

In a credit risk context, this thesis supports previous research that advanced ML algorithms are superior to traditional ones when used for financial prediction (Barboza et al., 2017; Khandani et al., 2010; Gu et al., 2020). The thesis shows that assessing future credit risk can be done more reliably when using advanced ML algorithms. The trained ML models are ready to use to make predictions of the credit risk of any firm for which relevant data are available. The result points to the expanding importance of machine learning in the paradigmatic fintech industry. This thesis further contributes to prior literature on CSP and credit risk (e.g., Kölbel et al., 2017; Jiraporn et al., 2014; Oikonomou et al., 2014) by testing to what extent ESG risks are incorporated in credit risk analysis of major credit rating agencies. In contrast to the finding of increased CDS spreads due to higher ESG risks (Kölbel et al., 2017), this thesis does not find that higher ESG risks are systematically reflected in credit ratings. Therefore, it indicates that credit ratings do not (yet) fully reflect the environmental and societal business conduct risks

that firms are facing. For instance, if a firm faces severe concerns in product safety, this is most likely not reflected in credit ratings, even though it might incur considerable cost (e.g., compensation payments) and reputational damages to the firm.

Overall, this thesis provides several new insights on CSP and the relationship between CSP and CFP. Therefore, it contributes to a variety of recent empirical and conceptual studies on potential drivers of the CSP-CFP relationship (e.g., Khan et al., 2016; Cai et al., 2016; Hong et al., 2019; Kölbel et al., 2017). Table 19 summarizes and compares the results and contributions of this thesis.

Table 19: Main Findings and Contributions

Paper	Main Findings	Main Contributions
First	Material CSP is positively related to the accruals quality proxy. Material and non-material CSP is negatively related to the real earnings management proxy. The Materiality of CSP does not significantly add explanatory power to the earnings quality when controlling for industry regulation.	Provides insights into the materiality of CSP. Develops an operationalization of material and non-material CSP using guidance from SASB. Uses data from Thomson Reuters and an extensive European sample with 542 firms from 27 different industries.
Second	Lower negative CSP is linked to a lower probability of class action lawsuits. Higher positive CSP protects against losses in market value during litigation. Implementing a trading strategy based on the findings yields monthly alphas of 1.31%.	Shows that running a business responsibly can reduce the number of lawsuits and thus prevent reputational damage and higher cost of capital. Predicts litigation risk as a function of CSP using actual class action lawsuits events. Calculates the actual market impact of litigation on firm value. Provides a portfolio approach that investors can use in a profitable way.
Third	Advanced ML algorithms considerably increase prediction accuracy for credit risk due to accounting for non-linearity and complex interaction effects. Considering information on negative CSP does not unambiguously increase prediction accuracy for credit risk.	Synthesizes credit risk prediction with state-of-the-art machine learning algorithms using credit rating data. The ready-to-use models allow market participants to assess future credit risk more reliably for any firm for which relevant data for the predictors are available. The analysis suggests that ESG risks are not (yet) systematically reflected in the credit risk analysis of major rating agencies.

Notes: This table describes the main findings and contributions to the seminal literature. Please note that for reasons of clarity the table is not exhaustive. More details on the specific findings and contributions of the papers can be found in Chapters II, III, and IV.

V.2 Note on a More General Limitation of Studies on CSP

This section discusses one of the more *general* limitations, which affects all three papers of this thesis but to varying extents. The rather *specific* limitations to the papers' research objectives and results are discussed under the respective sections (cf., Chapters II, III, and IV).

The most frequent criticism of the concept of CSP is often expressed in connection with its measurement. So-called ESG rating agencies advertise their scores on ESG performance and risks—like those primarily used in this thesis—as accurate and objective. A plethora of studies has widely and directly used these scores as they are supposed to measure a firm's CSP systematically and reliably. Many of these studies have already been cited in this thesis. Some research has used the scores more indirectly, i.e., by employing the membership in so-called sustainability indices as a proxy (e.g., Hawn et al., 2018; Belghitar et al., 2014). Examples of these indices include the Dow Jones Sustainability Index World or the FTSE4Good index series. However, ESG scores are used for the construction of these indices and are therefore the basis for any estimation of a firm's environmental, social, and governance performance.

Nevertheless, one rather crucial question is often left out, namely: Are these scores really as objective and reliable as the agencies claim? A couple of studies have tried to examine precisely this question empirically (Chatterji et al., 2009, 2016; Escrig-Olmedo et al., 2014, 2019).[71] Chatterji et al. (2009) tested the ratings of KLD—the most widely used rating source at the time. They showed that past environmental performance is adequately captured by environmental ratings. By comparing current ratings with actual future environmental performance, they further found that the negative ratings (i.e., those measuring environmental *concerns*) provide reasonable guidance on environmental performance. However, this finding did *not* hold for the positive ratings, i.e., those measuring environmental *strengths* (Chatterji et al., 2009). Even though this thesis does not use KLD ratings, these findings already suggest that ESG ratings do not necessarily reflect true performance, and measuring CSP is not a trivial endeavor. This

[71] The endeavor to investigate ESG ratings is impeded by the fact that the agencies' data collection process and calculation procedure (i.e., the algorithms) are proprietary to a large extent. This reticence makes it harder for researchers to get "a look behind the scenes." For example, the only information Thomson Reuters provides about its ratings is disclosed in a short PDF document, available via: https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/esg-scores-methodology.pdf. Similarly, RepRisk provides a PDF document, available via (login required): https://wrds-www.wharton.upenn.edu/documents/1016/RepRisk_-_Guidance_on_data_packages_and_elements_2019_version.pdf

conjecture becomes more clear when comparing the ratings of different agencies, including the ones that have been used in this thesis.

Chatterji et al. (2016) investigated the convergence of agencies by comparing their ratings. Surprisingly, they found a lack of consensus on the ESG ratings of six well-established agencies. After accounting for differences in the definition of CSR, they detected low validity of ESG ratings between agencies. Besides this lack of consensus, research suggested that aggregate measures are likely accompanied by the so-called "offset effect," i.e., positive and negative assessments offset each other (Escrig-Olmedo et al., 2014).[72] However, one should always keep in mind that many rating agencies were still in their infancy not too long ago and that they are working on improving their rating process continuously. For instance, a more recent study concluded that rating agencies added new criteria and improved their rating process to become more accurate and robust over the last ten years (Escrig-Olmedo et al., 2019). In conclusion, it can be said that while some of the problems associated with ratings can be circumvented (e.g., the offset effect), ESG ratings—like those used in this thesis—should at best be seen as an *approximation* of "true" performance.

This *approximation* leads to some caveats, some of which can be mitigated or even completely eliminated, and others, which must be regarded as limitations. On the one hand, the "offset effect" can be mitigated or overcome by using non-aggregated measures such as scores for each dimension of ESG and for negative CSP and positive CSP. Both approaches have been used in this thesis in various contexts. On the other hand, it is still difficult to judge the reliability and objectivity of available CSP measures, and there is little anyone can do about it other than empirical hypothesis testing using different sources of data. This caveat—among others—leads to possible avenues for future research.

V.3 Avenues for Future Research

For almost two decades, KLD has been perceived as "the de facto research standard" (Waddock, 2003, p. 369) for CSP. KLD is still a preferred source of ESG ratings for many researchers due

[72] The thesis at hand largely accounts for such a possible "offset effect" in one way or another. Please see the methodology sections under Chapters II, III, and IV for more details.

to the high acceptance that the ratings enjoy in the academic community. The caveats that come with this status have already been described in the previous section. However, since the beginning of the twenty-first century, new and innovative rating agencies have developed (e.g., RepRisk, Arabesque S-Ray, Climate Delta). These new agencies increasingly leverage the power of data science (e.g., artificial intelligence and big data) to design their ratings. With the emergence of new (and potentially more reliable) data on CSP, future research could unravel some of the long-standing questions that have never been conclusively answered. For example, there is still a lot about the responses of stakeholders to CSR initiatives or the temporal dimension of investments in CSR that research does not know—yet. The long-term nature of investments in CSR still needs more testing. Therefore, research needs to better understand the role of managerial myopia for the success of CSR investments. Gradually more reliable data on CSP alone may not provide the answers to all questions, but might be part of the solution to some problems.

Another topic that increasingly raises interest among researchers and corporations is the role that CSP plays for risk. Countless previous studies have focused on a *direct* and *positive* effect of CSP on the financial performance or firm value—many of them were also cited in this thesis. However, comparably few studies have considered the effect of CSP on the risk dimensions of a firm. While this thesis has tried to provide some answers regarding the effects on certain risk dimensions (e.g., litigation risk and credit risk), others remain un- or at least hardly explored. For example, the environmental risk dimension is likely to become significantly more pressing (cf., the Paris Agreement, or Fridays for Future). Early research in the 1990s focused on environmental performance (e.g., Russo and Fouts, 1997; Klassen and McLaughlin, 1996) but neglected the environmental risks that firms are facing. A potential explanation for the lack of attention is that most environmental risks (and some social risks) are still externalized to society. As soon as firms are forced to internalize these risks—either due to regulation or public/investor pressure or both—the business environment may change completely for some firms and industries. Only little research exists that examines what such a development would mean for market participants and society. For example, future research could investigate the effects of internalized risks on investor returns in certain regulatory environments (cf., “Konzernverantwortungsinitiative” in Switzerland).

There are many reasons suggesting that corporate environmental and social performance remain crucial—if they do not become even more vital—for both investors and society at large. At the beginning of the thesis, I mentioned the Fridays for Future movement and international agreements like the Paris Agreement, which both pursue mainly political goals. Other initiatives, like the so-called divestment movement "350," call for a major and eventually complete divestment in fossil fuels.[73] Divestment is another topic that is very likely to concern both politicians and market participants for quite some time in the near future—be it due to societal pressure (politicians) or risk mitigation (market participants). At this point, it seems reasonable to conclude that research opportunities concerning corporate environmental and social issues are getting more and more momentum.

[73] "350" refers to 350 parts per million, which is considered the upper limit for carbon dioxide concentration by scientific experts.

References

Aerts, W., Cormier, D., and Magnan, M. (2008). Corporate Environmental Disclosure, Financial Markets and the Media: An International Perspective. *Ecological Economics*, 64(3):643–659.

Agarwal, V. and Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking and Finance*, 32(8):1541–1551.

Aharony, J., Liu, C., and Yawson, A. (2015). Corporate litigation and executive turnover. *Journal of Corporate Finance*, 34:268–292.

Ajinkya, B., Bhojraj, S., and Sengupta, P. (2005). The association between outside directors, institutional investors and the properties of management earnings forecasts. *Journal of Accounting Research*, 43(3):343–376.

Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., and Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, 94:164–184.

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4):589–609.

Altman, E. I., Marco, G., and Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking & Finance*, 18(3):505–529.

Amel-Zadeh, A. and Serafeim, G. (2018). Why and how investors use ESG information: Evidence from a global survey. *Financial Analysts Journal*, 74(3):87–103.

Ang, J. and Mauck, N. (2011). Fire sale acquisitions: Myth vs. reality. *Journal of Banking & Finance*, 35(3):532–543.

Aouadi, A. and Marsat, S. (2018). Do ESG Controversies Matter for Firm Value? Evidence from International Data. *Journal of Business Ethics*, 151(4):1027–1047.

Arena, M. and Julio, B. (2015). The Effects of Securities Class Action Litigation on Corporate Liquidity and Investment Policy. *Journal of Financial and Quantitative Analysis*, 50(1-2):251–275.

Arena, M. P. (2018). Corporate litigation and debt. *Journal of Banking & Finance*, 87:202–215.

Ashbaugh-Skaife, H., Collins, D. W., and LaFond, R. (2006). The effects of corporate governance on firms' credit ratings. *Journal of Accounting and Economics*, 42(1-2):203–243.

Attig, N., Ghoul, S. E., Guedhami, O., and Suh, J. (2013). Corporate Social Responsibility and Credit Ratings. *Journal of Business Ethics*, 117(4):679–694.

Autore, D. M., Hutton, I., Peterson, D. R., and Smith, A. H. (2014). The effect of securities litigation on external financing. *Journal of Corporate Finance*, 27:231–250.

Avramov, D., Chordia, T., Jostova, G., and Philipov, A. (2009). Credit ratings and the cross-section of stock returns. *Journal of Financial Markets*, 12(3):469–499.

Baird, P. L., Geylani, P. C., and Roberts, J. A. (2012). Corporate Social and Financial Performance Re-Examined: Industry Effects in a Linear Mixed Model Analysis. *Journal of Business Ethics*, 109(3):367–388.

Barboza, F., Kimura, H., and Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83:405–417.

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

Basheer, I. A. and Hajmeer, M. (2000). Artificial neural networks: Fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43(1):3–31.

Bauer, R. and Braun, R. (2010). Misdeeds Matter: Long-Term Stock Price Performance after the Filing of Class-Action Lawsuits. *Financial Analysts Journal*, 66(6):74–92.

Belghitar, Y., Clark, E., and Deshmukh, N. (2014). Does it pay to be ethical? Evidence from the FTSE4Good. *Journal of Banking and Finance*, 47(1):54–62.

Benlemlih, M., Shaukat, A., Qiu, Y., and Trojanowski, G. (2018). Environmental and Social Disclosures and Firm Risk. *Journal of Business Ethics*, 152(3):613–626.

Bennell, J. A., Crabbe, D., Thomas, S., and Gwilym, O. A. (2006). Modelling sovereign credit ratings: Neural networks versus ordered probit. *Expert Systems with Applications*, 30(3):415–425.

Bhandari, L. C. (1988). Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence. *The Journal of Finance*, 43(2):507–528.

Bharath, S. T. and Shumway, T. (2008). Forecasting Default with the Merton Distance to Default Model. *The Review of Financial Studies*, 21(3):1339–1369.

Boehmer, E., Masumeci, J., and Poulsen, A. B. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics*, 30(2):253–272.

Boritz, J. E. and Kennedy, D. B. (1995). Effectiveness of neural network types for prediction of business failure. *Expert Systems With Applications*, 9(4):503–512.

Bose, I. and Mahapatra, R. K. (2001). Business data mining - A machine learning perspective. *Information and Management*, 39(3):211–225.

Bouslah, K., Kryzanowski, L., and M'Zali, B. (2013). The impact of the dimensions of social performance on firm risk. *Journal of Banking and Finance*, 37(4):1258–1273.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and regression trees*. Wadsworth International Group, Monterey, CA.

Breuer, W., Müller, T., Rosenbach, D., and Salzmann, A. (2018). Corporate social responsibility, investor protection, and cost of equity: A cross-country comparison. *Journal of Banking and Finance*, 96:34–55.

Brown, S., Hillegeist, S. A., and Lo, K. (2005). Management Forecasts and Litigation Risk.

Cai, L., Cui, J., and Jo, H. (2016). Corporate Environmental Responsibility and Firm Risk. *Journal of Business Ethics*, 139(3):563–594.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust Inference With Multiway Clustering. *Journal of Business and Economic Statistics*, 29(2):238–249.

Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008). In Search of Distress Risk. *The Journal of Finance*, 63(6):2899–2939.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1):57–82.

Carroll, A. B. (1979). A Three-Dimensional Conceptual Model of Corporate Performance. *The Academy of Management Review*, 4(4):497–505.

Carroll, A. B. (1991). The Pyramid of Corporate Social Responsibility: Toward the Moral Management of Organizational Stakeholders. *Business Horizons*, 34:39–48.

Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27:1–27:27.

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

Chatterji, A. K., Levine, D. I., and Toffel, M. W. (2009). How Well Do Social Ratings Actually Measure Corporate Social Responsibility? *Journal of Economics & Management Strategy*, 18(1):125–169.

Chava, S. (2014). Environmental externalities and cost of capital. *Management Science*, 60(9):2223–2247.

Chen, L. (2016). Local Institutions, Audit Quality, and Corporate Scandals of US-Listed Foreign Firms. *Journal of Business Ethics*, 133(2):351–373.

Chen, M.-Y. (2011). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers & Mathematics with Applications*, 62(12):4514–4524.

Chih, H.-L., Shen, C.-H., and Kang, F.-C. (2008). Corporate Social Responsibility, Investor Protection, and Earnings Management: Some International Evidence. *Journal of Business Ethics*, 79(1-2):179–198.

Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance*, 20(2):249–272.

Cleofas-Sánchez, L., García, V., Marqués, A. I., and Sánchez, J. S. (2016). Financial distress prediction using the hybrid associative memory with translation. *Applied Soft Computing Journal*, 44:144–152.

Cohen, D. A., Dey, A., and Lys, T. Z. (2008). Real and Accrual-Based Earnings Management in the Pre- and Post-Sarbanes-Oxley Periods. *The Accounting Review*, 83(3):757–787.

Cohen, D. A. and Zarowin, P. (2010). Accrual-based and real earnings management activities around seasoned equity offerings. *Journal of Accounting and Economics*, 50(1):2–19.

Cui, J., Jo, H., and Na, H. (2018). Does Corporate Social Responsibility Affect Information Asymmetry? *Journal of Business Ethics*, 148(3):549–572.

Dechow, P. M. and Dichev, I. D. (2002). The quality of accruals and earnings: The role of accrual estimation errors. *Accounting Review*, 77:35–59.

Dechow, P. M., Ge, W., and Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, 50(2-3):344–401.

Dechow, P. M., Sloan, R. G., and Sweeney, A. P. (1995). Detecting Earnings Management. *The Accounting Review*, 70(2):193–225.

Deng, S., Willis, R. H., and Xu, L. (2014). Shareholder Litigation, Reputational Loss, and Bank Loan Contracting. *Journal of Financial and Quantitative Analysis*, 49(4):1101–1132.

Derwall, J., Koedijk, K., and Ter Horst, J. (2011). A tale of values-driven and profit-seeking social investors. *Journal of Banking and Finance*, 35(8):2137–2147.

Dichev, I. D. (1998). Is the Risk of Bankruptcy a Systematic Risk? *The Journal of Finance*, 53(3):1131–1147.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263.

Dienes, D. and Velte, P. (2014). Einfluss der Corporate Governance auf die Nachhaltigkeitsberichterstattung. *Zeitschrift für Umweltpolitik und Umweltrecht/Journal of Environmental Law and Policy*, 37(4):347–389.

Dorfleitner, G., Grebler, J., and Utz, S. (2020). The Impact of Corporate Social and Environmental Performance on Credit Rating Prediction: North America versus Europe. *Journal of Risk, forthcoming*.

Drago, D., Carnevale, C., and Gallo, R. (2019). Do corporate social responsibility ratings affect credit default swap spreads? *Corporate Social Responsibility and Environmental Management*, 26(3):644–652.

Drucker, H. (1997). Improving Regressors using Boosting Techniques. Technical report, Monmouth University.

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

Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, 101(3):621–640.

El Ghoul, S., Guedhami, O., Kwok, C. C., and Mishra, D. R. (2011). Does corporate social responsibility affect the cost of capital? *Journal of Banking and Finance*, 35(9):2388–2406.

Ericsson, J. and Renault, O. (2006). Liquidity and credit risk. *The Journal of Finance*, 61(5):2219–2250.

Escrig-Olmedo, E., Fernández-Izquierdo, M. A., Ferrero-Ferrero, I., Rivera-Lirio, J. M., and Muñoz-Torres, M. J. (2019). Rating the raters: Evaluating how ESG rating agencies integrate sustainability principles. *Sustainability*, 11(3):915.

Escrig-Olmedo, E., Muñoz-Torres, M. J., Fernández-Izquierdo, M. A., and Rivera-Lirio, J. M. (2014). Lights and shadows on sustainability rating scoring. *Review of Managerial Science*, 8(4):559–574.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. (2008). LIBLINEAR: A Library for Large Linear Classification. *The Journal of Machine Learning Research*, 9:1871–1874.

Feng, Z.-Y., Wang, M.-L., and Huang, H.-W. (2015). Equity Financing and Social Responsibility: Further International Evidence. *The International Journal of Accounting*, 50(3):247–280.

Ferrell, A., Liang, H., and Renneboog, L. (2016). Socially Responsible Firms. *Journal of Financial Economics*, 122(3):585–606.

Flammer, C. (2013). Corporate Social Responsibility and Shareholder Reaction: The Environmental Awareness of Investors. *Academy of Management Journal*, 56(3):758–781.

Flammer, C. (2015). Does Corporate Social Responsibility Lead to Superior Financial Performance? A Regression Discontinuity Approach. *Management Science*, 61(11):2549–2568.

Flammer, C., Hong, B., and Minor, D. (2019). Corporate governance and the rise of integrating corporate social responsibility criteria in executive compensation: Effectiveness and implications for firm outcomes. *Strategic Management Journal*, 40(7):1097–1122.

Fombrun, C. J. and Shanley, M. (1990). What's in a Name? Reputation Building and Corporate Strategy. *Academy of Management Journal*, 33(2):233–258.

Francis, J., LaFond, R., Olsson, P., and Schipper, K. (2005). The market pricing of accruals quality. *Journal of Accounting and Economics*, 39(2):295–327.

Francis, J., Philbrick, D., and Schipper, K. (1994). Shareholder Litigation and Corporate Disclosures. *Journal of Accounting Research*, 32(2):137–164.

Frazzini, A., Israel, R., and Moskowitz, T. J. (2014). Trading Costs of Asset Pricing Anomalies.

Freeman, R. E. (2010). *Strategic Management: A Stakeholder Approach*. Cambridge University Press, Cambridge.

Freund, Y. and Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1):119–139.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232.

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

Fritzsche, D. J. (1991). A model of decision-making incorporating ethical values. *Journal of Business Ethics*, 10(11):841–852.

Fu, F. (2009). Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics*, 91(1):24–37.

Gande, A. and Lewis, C. M. (2009). Shareholder-Initiated Class Action Lawsuits: Shareholder Wealth Effects and Industry Spillovers. *The Journal of Financial and Quantitative Analysis*, 44(4):823–850.

Gao, L. and Zhang, J. H. (2015). Firms' earnings smoothing, corporate social responsibility, and valuation. *Journal of Corporate Finance*, 32:108–127.

Garriga, E. and Melé, D. (2004). Corporate Social Responsibility Theories: Mapping the Territory. *Journal of Business Ethics*, 53(1/2):51–71.

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., Sebastopol, CA, 1st edition.

Godfrey, P. C. (2005). The Relationship Between Corporate Philanthropy And Shareholder Wealth: A Risk Management Perspective. *Academy of Management Review*, 30(4):777–798.

Godfrey, P. C., Merrill, C. B., and Hansen, J. M. (2009). The relationship between corporate social responsibility and shareholder value: an empirical test of the risk management hypothesis. *Strategic Management Journal*, 30(4):425–445.

Goss, A. and Roberts, G. S. (2011). The impact of corporate social responsibility on the cost of bank loans. *Journal of Banking and Finance*, 35(7):1794–1810.

Gow, I. D., Ormazabal, G., and Taylor, D. J. (2010). Correcting for cross-cectional and time-series dependence in accounting research. *The Accounting Review*, 85(2):483–512.

Graham, J. R. and Harvey, C. R. (2001). The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics*, 60(2-3):187–243.

Grewatsch, S. and Kleindienst, I. (2017). When Does It Pay to be Good? Moderators and Mediators in the Corporate Sustainability–Corporate Financial Performance Relationship: A Critical Review. *Journal of Business Ethics*, 145(2):383–416.

Grüner, A., Gutsche, R., and Schulz, J.-F. (2014). The Value-Relevance of CSR Reporting Quality. *Zeitschrift für Umweltpolitik und Umweltrecht/Journal of Environmental Law and Policy*, 37(4):390–409.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33(5):2223–2273.

Gupta, K. (2018). Environmental Sustainability and Implied Cost of Equity: International Evidence. *Journal of Business Ethics*, 147(2):343–365.

Harjoto, M. and Laksmana, I. (2018). The Impact of Corporate Social Responsibility on Risk Taking and Firm Value. *Journal of Business Ethics*, 151(2):353–373.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Haslem, B. (2005). Managerial Opportunism during Corporate Litigation. *The Journal of Finance*, 60(4):2013–2041.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer-Verlag, New York, 2nd edition.

Hawn, O., Chatterji, A. K., and Mitchell, W. (2018). Do investors actually value sustainability? New evidence from investor reactions to the Dow Jones Sustainability Index (DJSI). *Strategic Management Journal*, 39(4):949–976.

Hegarty, W. H. and Sims, H. P. (1978). Some determinants of unethical decision behavior: An experiment. *Journal of Applied Psychology*, 63(4):451–457.

Hemingway, C. A. and Maclagan, P. W. (2004). Managers' Personal Values as Drivers of Corporate Social Responsibility. *Journal of Business Ethics*, 50(1):33–44.

Henisz, W. J. and McGlinch, J. (2019). ESG, Material Credit Events, and Credit Risk. *Journal of Applied Corporate Finance*, 31(2):105–117.

Hinze, A.-K. and Freidank, C.-C. (2018). Kritische Analyse moglicher Ausgestaltungen der nichtfinanziellen Erklärung vor dem Hintergrund des CSR-Richtlinie-Umsetzungsgesetzes. *Zeitschrift für Umweltpolitik und Umweltrecht/Journal of Environmental Law and Policy*, 41(1):21–51.

Hoerl, A. E. (1962). Application of Ridge Analysis to Regression Problems. *Chemical Engineering Progress*, 58(3):54–59.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55–67.

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

Hong, H., Kubik, J. D., Liskovich, I., and Scheinkman, J. (2019). Crime, Punishment and the Value of Corporate Social Responsibility. *SSRN Electronic Journal*.

Hong, Y. and Andersen, M. L. (2011). The Relationship Between Corporate Social Responsibility and Earnings Management: An Exploratory Study. *Journal of Business Ethics*, 104(4):461–471.

Houweling, P. and Van Zundert, J. (2017). Factor investing in the corporate bond market. *Financial Analysts Journal*, 73(2):100–115.

Hsu, F. J. and Chen, Y.-C. (2015). Is a firm's financial risk associated with corporate social responsibility? *Management Decision*, 53(9):2175–2199.

Humphrey, J. E., Lee, D. D., and Shen, Y. (2012). Does it cost to be sustainable? *Journal of Corporate Finance*, 18(3):626–639.

Ibikunle, G. and Steffen, T. (2017). European Green Mutual Fund Performance: A Comparative Analysis with their Conventional and Black Peers. *Journal of Business Ethics*, 145(2):337–355.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., and Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2(4):230–243.

Jiraporn, P., Jiraporn, N., Boeprasert, A., and Chang, K. (2014). Does corporate social responsibility (CSR) improve credit ratings? Evidence from geographic identification. *Financial Management*, 43(3):505–531.

Jo, H., Kim, H., and Park, K. (2015). Corporate Environmental Responsibility and Firm Performance in the Financial Services Sector. *Journal of Business Ethics*, 131(2):257–284.

Jo, H. and Na, H. (2012). Does CSR Reduce Firm Risk? Evidence from Controversial Industry Sectors. *Journal of Business Ethics*, 110(4):441–456.

Johnson, M. F., Kasznik, R., and Nelson, K. K. (2000). Shareholder Wealth Effects of the Private Securities Litigation Reform Act of 1995. *Review of Accounting Studies*, 5(3):217–233.

Johnson, M. F., Kasznik, R., and Nelson, K. K. (2001). The Impact of Securities Litigation Reform on the Disclosure of Forward-Looking Information by High Technology Firms. *Journal of Accounting Research*, 39(2):297–327.

Jones, J. J. (1991). Earnings Management During Import Relief Investigations. *Journal of Accounting Research*, 29(2):193–228.

Jordaan, L. A., de Klerk, M., and de Villiers, C. J. (2018). Corporate social responsibility and earnings management of South African companies. *South African Journal of Economic and Management Sciences*, 21(1).

Karpoff, J. M., Lee, D. S., and Martin, G. S. (2008). The Cost to Firms of Cooking the Books. *The Journal of Financial and Quantitative Analysis*, 43(3):581–611.

Kealhofer, S. (2003). Quantifying credit risk I: Default prediction. *Financial Analysts Journal*, 59(1):30–44.

Kennett, D. A. (1980). Altruism and Economic Behavior: II Private Charity and Public Policy. *American Journal of Economics and Sociology*, 39(4):337–352.

Khan, M. (2019). Corporate Governance, ESG, and Stock Returns around the World. *Financial Analysts Journal*, 75(4):103–123.

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

Khandani, A. E., Kim, A. J., and Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking and Finance*, 34(11):2767–2787.

Kiesel, F. and Lücke, F. (2019). ESG in credit ratings and the impact on financial markets. *Financial Markets, Institutions & Instruments*, 28(3):263–290.

Kiessling, T., Isaksson, L., and Yasar, B. (2015). Market Orientation and CSR: Performance Implications. *Journal of Business Ethics*, 137(2):269–284.

Kim, I. and Skinner, D. J. (2012). Measuring securities litigation risk. *Journal of Accounting and Economics*, 53(1-2):290–310.

Kim, I. and Venkatachalam, M. (2011). Are Sin Stocks Paying the Price for Accounting Sins? *Journal of Accounting, Auditing & Finance*, 26(2):415–442.

Kim, Y., Li, H., and Li, S. (2014). Corporate Social Responsibility and Stock Price Crash Risk. *Journal of Banking and Finance*, 43(1):1–13.

Kim, Y. and Park, M. S. (2005). Pricing of Seasoned Equity Offers and Earnings Management. *Journal of Financial and Quantitative Analysis*, 40(2):435–463.

Kim, Y., Park, M. S., and Wier, B. (2012). Is Earnings Quality Associated with Corporate Social Responsibility? *The Accounting Review*, 87(3):761–796.

Kingma, D. P. and Ba, J. L. (2015). Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR.

Kisgen, D. J. (2006). Credit ratings and capital structure. *Journal of Finance*, 61(3):1035–1072.

Klassen, R. D. and McLaughlin, C. P. (1996). The impact of environmental management on firm performance. *Management Science*, 42(8):1199–1214.

Koh, P. S., Qian, C., and Wang, H. (2014). Firm litigation risk and the insurance value of corporate social performance. *Strategic Management Journal*, 35(10):1464–1482.

Kölbel, J. F., Busch, T., and Jancso, L. M. (2017). How Media Coverage of Corporate Social Irresponsibility Increases Financial Risk. *Strategic Management Journal*, 38(11):2266–2284.

Kothari, S. P., Leone, A. J., and Wasley, C. E. (2005). Performance matched discretionary accrual measures. *Journal of Accounting and Economics*, 39(1):163–197.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., and Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13:8–17.

Krüger, P. (2015). Corporate Goodness and Shareholder Wealth. *Journal of Financial Economics*, 115(2):304–329.

Labelle, R., Makni Gargouri, R., and Francoeur, C. (2010). Ethics, Diversity Management, and Financial Reporting Quality. *Journal of Business Ethics*, 93(2):335–353.

Lee, D. D. and Faff, R. W. (2009). Corporate Sustainability Performance and Idiosyncratic Risk: A Global Perspective. *Financial Review*, 44(2):213–237.

Leland, H. E. (1994). Corporate Debt Value, Bond Covenants, and Optimal Capital Structure. *The Journal of Finance*, 49(4):1213–1252.

Lev, B., Petrovits, C., and Radhakrishnan, S. (2010). Is Doing Good Good for You? How Corporate Charitable Contributions Enhance Revenue Growth. *Strategic Management Journal*, 31(2):182–200.

Liang, H. and Renneboog, L. (2017). On the Foundations of Corporate Social Responsibility. *The Journal of Finance*, 72(2):853–910.

Lins, K. V., Servaes, H., and Tamayo, A. (2017). Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis. *The Journal of Finance*, 72(4):1785–1824.

Liu, C., Cheong, C. S., and Zurbruegg, R. (2020). Rhetoric, Reality, and Reputation: Do CSR and Political Lobbying Protect Shareholder Wealth against Environmental Lawsuits? *Journal of Financial and Quantitative Analysis*, 55(2):679–706.

Lowry, M. and Shu, S. (2002). Litigation risk and IPO underpricing. *Journal of Financial Economics*, 65(3):309–335.

Mackey, A., Mackey, T. B., and Barney, J. B. (2007). Corporate Social Responsibility and Firm Performance: Investor Preferences and Corporate Strategies. *Academy of Management Review*, 32(3):817–835.

Makni, R., Francoeur, C., and Bellavance, F. (2009). Causality Between Corporate Social Performance and Financial Performance: Evidence from Canadian Firms. *Journal of Business Ethics*, 89(3):409–422.

Margolis, J. D., Elfenbein, H. A., and Walsh, J. P. (2009). Does it Pay to Be Good...And Does it Matter? A Meta-Analysis of the Relationship between Corporate Social and Financial Performance. *SSRN Electronic Journal*.

Martínez-Ferrero, J., Banerjee, S., and García-Sánchez, I. M. (2016). Corporate Social Responsibility as a Strategic Shield Against Costs of Earnings Management Practices. *Journal of Business Ethics*, 133(2):305–324.

Matsumoto, D. A. (2002). Management's Incentives to Avoid Negative Earnings Surprises. *The Accounting Review*, 77(3):483–514.

Matsumura, E. M., Prakash, R., and Vera-Muñoz, S. C. (2014). Firm-Value Effects of Carbon Emissions and Carbon Disclosures. *The Accounting Review*, 89(2):695–724.

McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(2):109–142.

McNichols, M. F. (2002). The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors: Discussion. *The Accounting Review*, 77(Suppl.):61–69.

McWilliams, A. and Siegel, D. (2000). Corporate social responsibility and financial performance: Correlation or misspecification? *Strategic Management Journal*, 21(5):603–609.

McWilliams, A., Siegel, D. S., and Wright, P. M. (2006). Corporate Social Responsibility: Strategic Implications. *Journal of Management Studies*, 43(1):1–18.

Menz, K.-M. M. (2010). Corporate Social Responsibility: Is it Rewarded by the Corporate Bond Market? A Critical Note. *Journal of Business Ethics*, 96(1):117–134.

Merton, R. C. (1973). Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183.

Merton, R. C. (1974). On the Pricing of Corporate Debt: the Risk Structure of Interest Rates. *The Journal of Finance*, 29(2):449–470.

Merton, R. C. (2006). A Simple Model of Capital Market Equilibrium with Incomplete Information. *The Journal of Finance*, 42(3):483.

Niu, T., Wang, J., Lu, H., Yang, W., and Du, P. (2020). Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting. *Expert Systems with Applications*, 148:1–17.

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1):109–131.

Oikonomou, I., Brooks, C., and Pavelin, S. (2014). The Effects of Corporate Social Performance on the Cost of Corporate Debt and Credit Ratings. *The Financial Review*, 49(1):49–75.

Orlitzky, M., Schmidt, F. L., and Rynes, S. L. (2003). Corporate Social and Financial Performance: A Meta-Analysis. *Organization Studies*, 24(3):403 –441.

Park, Y. W. and Shin, H. H. (2004). Board composition and earnings management in Canada. *Journal of Corporate Finance*, 10(3):431–457.

Piramuthu, S. (2006). On preprocessing data for financial credit risk evaluation. *Expert Systems with Applications*, 30(3):489–497.

PRI (2018). ESG in Credit Risk and Ratings Forums: Investor Survey Results. Technical report, PRI Association, London.

Prior, D., Surroca, J., and Tribó, J. A. (2008). Are Socially Responsible Managers Really Ethical? Exploring the Relationship Between Earnings Management and Corporate Social Responsibility. *Corporate Governance: An International Review*, 16(3):160–177.

Rasekhschaffe, K. C. and Jones, R. C. (2019). Machine Learning for Stock Selection. *Financial Analysts Journal*, 75(3):70–88.

Rogers, J. L. and Stocken, P. C. (2005). Credibility of Management Forecasts. *The Accounting Review*, 80(4):1233–1260.

Roychowdhury, S. (2006). Earnings management through real activities manipulation. *Journal of Accounting and Economics*, 42(3):335–370.

Ruder, S. (2017). An overview of gradient descent optimization algorithms. Technical report, Insight Centre for Data Analytics, NUI Galway.

Russo, M. V. and Fouts, P. A. (1997). A Resourced-Based Perspective on Corporate Environmental Performance and Profitability. *Academy of Management Journal*, 40(3):534–559.

Sassen, R., Hinze, A.-K., and Hardeck, I. (2016). Impact of ESG Factors on Firm Risk in Europe. *Journal of Business Economics*, 86(8):867–904.

Schnietz, K. E. and Epstein, M. J. (2005). Exploring the Financial Value of a Reputation for Corporate Social Responsibility During a Crisis. *Corporate Reputation Review*, 7(4):327–345.

Schreck, P. (2011). Reviewing the Business Case for Corporate Social Responsibility: New Evidence and Analysis. *Journal of Business Ethics*, 103(2):167–188.

Schwert, G. W. (1983). Size and stock returns, and other empirical regularities. *Journal of Financial Economics*, 12(1):3–12.

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1):1–47.

Shanker, M. S., Hu, M. Y., and Hung, M. S. (1996). Effect of data standardization on neural network training. *Omega*, 24(4):385–397.

Sharfman, M. P. and Fernando, C. S. (2008). Environmental risk management and the cost of capital. *Strategic Management Journal*, 29(6):569–592.

Smith, C. W. and Stulz, R. M. (1985). The Determinants of Firms' Hedging Policies. *Journal of Financial and Quantitative Analysis*, 20(4):391.

Stanwick, P. A. and Stanwick, S. D. (1998). The relationship between corporate social performance, and organizational size, financial performance, and environmental performance: An empirical examination. *Journal of Business Ethics*, 17(2):195–204.

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

Strike, V. M., Gao, J., and Bansal, P. (2006). Being good while being bad: Social responsibility and the international diversification of US firms. *Journal of International Business Studies*, 37(6):850–862.

Sun, W. and Cui, K. (2014). Linking corporate social responsibility to firm default risk. *European Management Journal*, 32(2):275–287.

Swanson, D. L. (1995). Addressing a Theoretical Problem by Reorienting the Corporate Social Performance Model. *Academy of Management Review*, 20(1):43–64.

Taylor, M. and Watts, J. (2019). Revealed: the 20 firms behind a third of all carbon emissions.

Teoh, S. H., Welch, I., and Wong, T. J. (1998). Earnings management and the long-run market performance of initial public offerings. *The Journal of Finance*, 53(6):1935–1974.

Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics*, 99(1):1–10.

Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3):475–494.

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

Vapnik, V. N. (1998). *Statistical learning theory*. New York: John Wiley & Sons.

Vassalou, M. and Xing, Y. (2004). Default Risk in Equity Returns. *The Journal of Finance*, 59(2):831–868.

Verwijmeren, P. and Derwall, J. (2010). Employee well-being, firm leverage, and bankruptcy risk. *Journal of Banking & Finance*, 34(5):956–964.

Waddock, S. (2003). Myths and Realities of Social Investing. *Organization & Environment*, 16(3):369–380.

Waddock, S. A. and Graves, S. B. (1997). The corporate social performance-financial performance link. *Strategic Management Journal*, 18(4):303–319.

Wartick, S. L. and Cochran, P. L. (1985). The Evolution of the Corporate Social Performance Model. *The Academy of Management Review*, 10(4):758–769.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., and et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):1–9.

Wood, D. J. (1991). Corporate Social Performance Revisited. *The Academy of Management Review*, 16(4):691–718.

Wu, L. and Yang, Y. (2014). Nonnegative Elastic Net and application in index tracking. *Applied Mathematics and Computation*, 227:541–552.

Yeh, C. C., Chi, D. J., and Lin, Y. R. (2014). Going-concern prediction using hybrid random forests and rough set approach. *Information Sciences*, 254:98–110.

Zang, A. Y. (2012). Evidence on the Trade-Off between Real Activities Manipulation and Accrual-Based Earnings Management. *The Accounting Review*, 87(2):675–703.

Zou, H. and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(2):301–320.

Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., and Telenti, A. (2019). A primer on deep learning in genomics. *Nature Genetics*, 51(1):12–18.

A Appendix: Paper 1

A.1 Materiality Guidance of the Sustainability Accounting Standards Board

The SASB was one of the first to develop guidance on the materiality of ESG investments. It is an independent (non-profit) standards-setting organization, which strives to enhance capital markets efficiency by developing guidance for the disclosure of material ESG information.[74] In general, SASB considers four crucial aspects in its standards-setting process: First, it takes on the perspective of an investor; second, it bases its guidance on evidence from research; third, it considers market aspects; and last, it focuses on materiality.

By March 2018, guidance from the SASB covered 79 industries across 11 sectors. The standards-setting process of SASB includes evidence-based research, quantitative data analysis, a public comment period, and the inclusion of various stakeholder groups. The Sector Advisory Groups are such stakeholders that comprise volunteering industry experts and advise the SASB in different stages of the standards development process. In addition, Industry Working Groups comment on the guidance developed by SASB in a later stage of the standards development process.

SASB explicitly focuses on a shareholder perspective and thus differs fundamentally from other standards-setting organizations. For instance, the Global Reporting Initiative develops sustainability reporting standards that follow a much broader, multi-stakeholder approach, which is rooted in the public interest.[75] The GRI also expresses this in its mission statement, which is "to empower decisions that create social, environmental and economic benefits for everyone."

The guidance on materiality by SASB determines whether a firm's performance on a given ESG topic has an influence on firm value and thus potentially affects investors' decisions. If information about ESG issues potentially affects an investor's investment (or voting) decision, it is likely to be material according to the SASB. More specifically, the SASB follows the

[74] More information on SASB can be found on https://www.sasb.org/

[75] More information on GRI via: https://www.globalreporting.org

definition of material information by the US Supreme Court. According to this definition, material information represents "a substantial likelihood that the disclosure of the omitted fact would have been viewed by the reasonable investor as having significantly altered the 'total mix' of information made available."[76]

Khan et al. (2016) provide a detailed explanation of SASB's standards development process and the guidance on materiality. Extensive information is also available via the SASB website (www.sasb.org). Therefore, I refrain from presenting every detail once again and invite the interested reader to consult the relevant sources. Nevertheless, I make use of the guidance on materiality provided by SASB.[77]

A.2 Yearly Sample Distribution

Table A.1: Number of Observations per Industry and Year

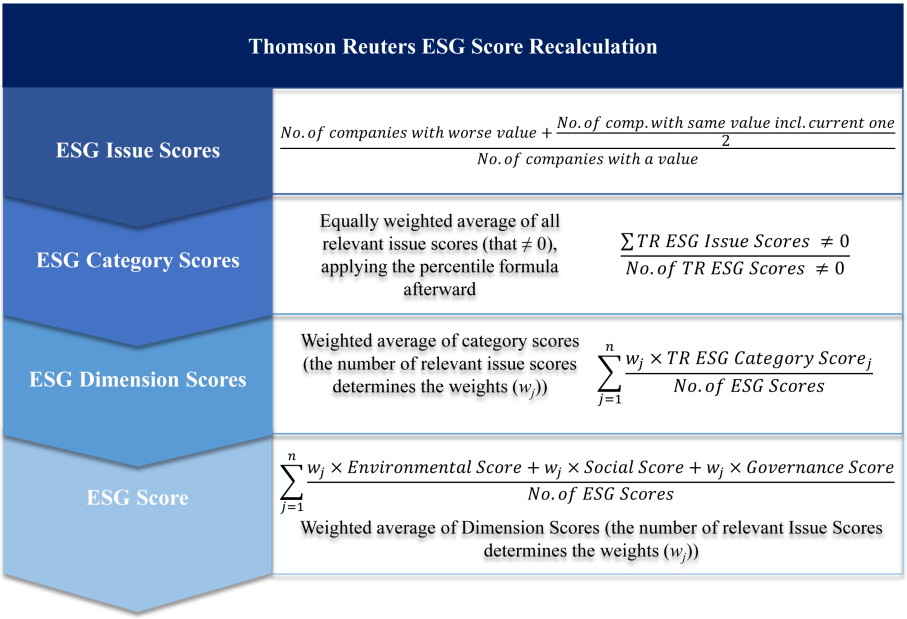
Year	# of Industries	# of Observations
2005	23	374
2006	25	403
2007	25	434
2008	26	473
2009	27	508
2010	27	536
2011	27	542
2012	27	542
2013	27	542
2014	27	540
2015	27	541
2016	27	539
Total		5,974

[76] TSC Industries v. Northway, 426, US 438, 449 (1976). Available via: https://supreme.justia.com/cases/federal/us/426/438/case.html.

[77] The SASB kindly provided the materiality data under an academic license agreement. © 2014 The SASB Foundation. All Rights Reserved.

A.3 Overview of the TR ESG Score Recalculation Procedure

Figure A.1: Step-by-Step Calculation of TR ESG Scores



A.4 Derivation of the Modified Accruals Quality Model

The modified accruals quality model (AQM) consists of three parts. First, it measures normal accruals by relating TCA to operating cash flows ($CFO_{j,t-1}$, $CFO_{j,t}$, and $CFO_{j,t+1}$). Second, it measures abnormal accruals by two variables from the modified Jones model (Dechow et al., 1995; Jones, 1991; McNichols, 2002): change in revenue (ΔREV) and PPE. Third, it includes prior-year ROA to control for the effect of performance on abnormal accruals (Kim et al., 2012; Kothari et al., 2005).

The modified AQM looks as follows:[78]

[78] All variables are scaled by lagged and averaged total assets (TA), except for ROA. By nature, ROA is already a scaled variable.

$$\frac{TCA_{j,t}}{TA_{j,t-1}} = \alpha_0 + \beta_1 \frac{CFO_{j,t-1}}{TA_{j,t-2}} + \beta_2 \frac{CFO_{j,t}}{TA_{j,t-1}} + \beta_3 \frac{CFO_{j,t+1}}{TA_{j,t}} + \beta_4 \frac{\Delta REV_{j,t}}{TA_{j,t-1}} + \beta_5 \frac{PPE_{j,t}}{TA_{j,t-1}} + \beta_6 ROA_{j,t-1} + \varepsilon_{j,t} \tag{A.4.1}$$

where:

$TCA_{j,t}$ is firm j's total current accruals, calculated as $\Delta CA_{j,t} - \Delta CL_{j,t} - \Delta Cash_{j,t} + \Delta STDEBT_{j,t}$

$TA_{j,t}$ is a firm j's total assets in year t

$\Delta \text{CA}_{j,t}$ is firm j's change in current assets between the last $(t-1)$ and the current (t) year

$\Delta \text{CL}_{j,t}$ is firm j's change in current liabilities between the last $(t-1)$ and the current (t) year

$\Delta Cash_{j,t}$ is firm j's change in cash between the last $(t-1)$ and the current (t) year

$\Delta \text{STDEBT}_{j,t}$ is firm j's change in debt in current liabilities between the last $(t-1)$ and the current (t) year

$CFO_{j,t-1}$ is firm j's operating cash flow in year $t-1$

$CFO_{j,t}$ is firm j's operating cash flow in year t

$CFO_{j,t+1}$ is firm j's operating cash flow in year $t+1$

$\Delta REV_{j,t}$ is firm j's change in revenue between the last $(t-1)$ and the current (t) year

$PPE_{j,t}$ is firm j's value of PPE in year t

$ROA_{j,t-1}$ is firm j's industry-adjusted return on assets in year $t-1$

The main dependent variable of interest, year- and firm-specific accruals quality (AQ), is calculated by the absolute yearly value of firm j's residuals ($|\varepsilon_{j,t}|$) from the cross-sectional

regressions of the modified AQM on industry-level including country fixed-effects (Kim et al., 2012; Kothari et al., 2005). Including country fixed-effects was necessary to account for differences in financial reporting and regulatory standards between the 24 European countries in which the sample firms of this study have their headquarter. The higher the $|\varepsilon_{j,t}|$, the lower the accruals quality and the more likely that firm j is subject to earnings management.

A.5 Derivation of the Real Earnings Management (REM) Models

This sub-section describes the three proxies for REM that were proposed and used by previous research (e.g., Roychowdhury, 2006; Cohen et al., 2008; Zang, 2012; Kim et al., 2012): ACFO, APROD, and ADIEXP. Therefore, this study models ACFO by taking the residuals ($\varepsilon_{j,t}$) from regressing sales and change in sales on the actual operating cash flow:[79]

$$\frac{CFO_{j,t}}{TA_{j,t-1}} = \alpha_0 + \alpha_1 \frac{1}{TA_{j,t-1}} + \beta_1 \frac{S_{j,t}}{TA_{j,t-1}} + \beta_2 \frac{\Delta S_{j,t}}{TA_{j,t-1}} + \varepsilon_{j,t} \tag{A.5.1}$$

where:

$CFO_{j,t}$ is a firm j's operating cash flow in year t

$TA_{j,t-1}$ is a firm j's total assets in year $t-1$

$S_{j,t}$ is a firm j's sales in year t

$\Delta S_{j,t}$ is a firm j's change in sales from year $t-1$ to year t

APROD was considered as another proxy for earnings management. Following previous research (e.g., Kim et al., 2012), this study defined production costs as the sum of cost of goods sold (COGS) in year t and change in inventory (ΔINV) in year t. Abnormal production costs were estimated by running the following industry-year regression and taking the residuals from the model ($\varepsilon_{j,t}$) as the appropriate measure for APROD (Roychowdhury, 2006).

[79] All variables are scaled by lagged TA. In line with Roychowdhury (2006), the models include a scaled intercept ($\alpha_1(\frac{1}{TA_{j,t-1}})$). This inclusion is necessary to avoid spurious correlation between the scaled dependent and independent variables due to variation in the scaling variable (i.e., total assets).

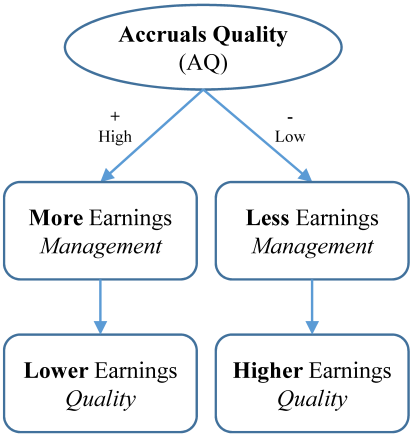
$$\frac{PROD_{j,t}}{TA_{j,t-1}} = \alpha_0 + \alpha_1 \frac{1}{TA_{j,t-1}} + \beta_1 \frac{S_{j,t}}{TA_{j,t-1}} + \beta_2 \frac{\Delta S_{j,t}}{TA_{j,t-1}} + \beta_3 \frac{S_{j,t-1}}{TA_{j,t-1}} + \varepsilon_{j,t} \quad \text{(A.5.2)}$$

Finally, ADIEXP was considered as another proxy for earnings management. Abnormal discretionary expense is calculated as follows (again, the residuals ($\varepsilon_{j,t}$) from the model measure abnormal discretionary expense):

$$\frac{DIEXP_{j,t}}{TA_{j,t-1}} = \alpha_0 + \alpha_1 \frac{1}{TA_{j,t-1}} + \beta_1 \frac{S_{j,t-1}}{TA_{j,t-1}} + \varepsilon_{j,t} \quad \text{(A.5.3)}$$

A.6 Interpretation of the Accruals Quality Proxy

Figure A.2: Interpreting the Sign of Accruals Quality



A.7 Interpretation of the Real Earnings Management Proxies

Figure A.3: Interpreting the Sign of Real Earnings Management

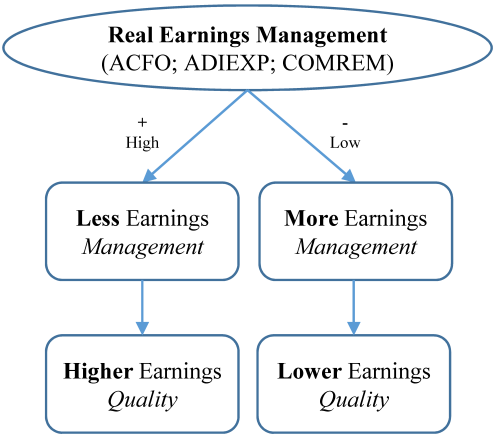
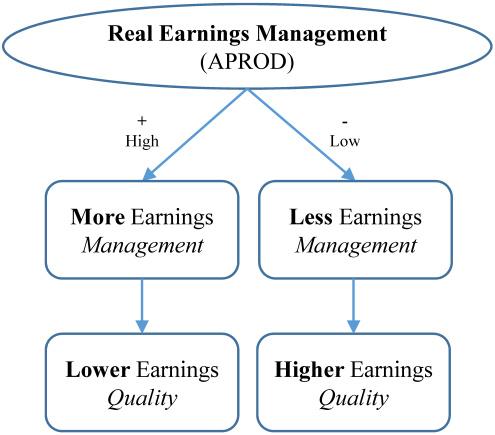


Figure A.4: Interpreting the Sign of Real Earnings Management (2)



A.8 Robustness: Additional Analysis using Positive and Negative AQ

Table A.2: Robustness Check – Regressing Positive and Negative AQ on Material and Non-Material CSP

	Positive AQ		Negative AQ	
Independent Variables	(1)	(2)	(3)	(4)
Material CSP	−0.023		0.050**	
	(−1.36)		(3.59)	
Non-Material CSP		−0.024		0.025
		(−1.27)		(1.41)
Control Variables				
COMBREM	0.099**	0.099**	0.031**	0.030**
	(7.90)	(7.99)	(2.96)	(2.87)
WorkingCapital	0.181**	0.181**	0.027	0.027
	(7.53)	(7.50)	(1.29)	(1.29)
RDActivities	0.573**	0.579**	−0.495**	−0.482**
	(5.35)	(5.39)	(−3.58)	(−3.43)
CGScore	0.005	0.005	−0.009	−0.008
	(0.36)	(0.36)	(−0.70)	(−0.65)
Age	0.000**	0.000**	0.000*	0.000*
	(−4.37)	(−4.35)	(2.06)	(2.19)
PriceEarnings	0.000	0.000	0.000	0.000
	(−0.35)	(−0.39)	(−0.61)	(−0.52)
MKTB	0.004**	0.004**	0.003*	0.003*
	(3.99)	(3.95)	(2.40)	(2.34)
ROA	0.390**	0.391**	0.099	0.107
	(6.82)	(6.84)	(1.80)	(1.94)
Leverage	0.022	0.022	−0.027	−0.030
	(0.96)	(0.98)	(−1.20)	(−1.31)
EOIndicator	−0.022**	−0.022**	0.009	0.009
	(−2.65)	(−2.68)	(1.55)	(1.49)
Constant	−0.105**	−0.103**	−0.228	−0.229
	(−3.42)	(−3.37)	(−1.87)	(−1.88)
Country FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	3462	3462	1886	1886
Adjusted R2	0.26	0.26	0.19	0.19

Notes: This table shows coefficients for the multiple regressions of Positive AQ and Negative AQ (i.e., the two sub-samples) on Material CSP and Non-Material CSP. All specifications include country and industry fixed-effects. For variable descriptions, see Table 4. t statistics in parentheses. * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

B Appendix: Paper 2

B.1 Variable Definitions

Table B.1: Variable Definitions

Variable	Type	Description
ClassPeriod	$\{0;1\}$	Takes the value of 1 if the firm is affected by a class period (i.e., the firm has been potentially involved in punishable endeavors at the time) and 0 otherwise.
CaseFiled	$\{0;1\}$	Takes the value of 1 if the firm is affected by a case filing (i.e., a class action lawsuit has been effectively filed against the firm) and 0 otherwise.
ClassPeriod Start	$\{0;1\}$	Takes the value of 1 if the firm is affected by the start of a class period (i.e., the firm has started to be potentially involved in punishable endeavors at the time) and 0 otherwise.
ESG_P-score	$[0;1]$	The Thomson Reuters methodology uses a percentile score formula that benchmarks each firm against its industry group peers for ten ESG categories. Aggregating the score for each category results in an overall ESG score. Higher values of this score thus indicate a higher ESG performance of a firm compared to its industry group peers.
ESG_N-score	$[0;1]$	The Thomson Reuters methodology uses a percentile score formula that benchmarks each firm against its industry group peers using 23 ESG controversy topics. This procedure yields one ESG controversy score, which reflects how strongly a firm has been involved in ESG controversies compared to its industry group peers.
ENV_P-score	$[0;1]$	Score that considers only environmental issues.
SOC_P-score	$[0;1]$	Score that considers only social issues.
GOV_P-score	$[0;1]$	Score that considers only governance issues.
FPS	$\{0;1\}$	Takes the value of 1 if the firm is a member of the biotechnology, information technology, or retail industry. These industries have historically the highest litigation rates.
Size	$[0;\infty]$	Natural logarithm of total assets.
SalesGrowth	$[-1;\infty]$	Change in sales compared to the previous year.
Ret	$[-1;\infty]$	Market-adjusted 12-month stock return.
RetSkew	$[-\infty;\infty]$	Skewness of the 12-month stock return.
RetStdDev	$[0;\infty]$	Standard deviation of the 12-month stock return.
Turnover	$[0;\infty]$	Total 12-month stock turnover scaled by average shares outstanding.
BTM	$[-\infty;\infty]$	End-of-fiscal-year shareholders equity divided by total market value of shares outstanding.
Leverage	$[-\infty;\infty]$	End-of-fiscal-year long-term debt divided by shareholders equity.
ROE	$[-\infty;\infty]$	Net income divided by shareholders equity.

B.2 Examples for Class Action Lawsuits

Table B.2: Detailed Information on Examples for Class Action Lawsuits

Firm	Year	Market	Case Status	Class Period Days	Summary
Apple Inc	2006	NASDAQ	Settled	576	The action alleges that defendants made false and misleading statements and omissions concerning Apple's improper and undisclosed practice of backdating options conferred on certain executives which made it appear that such options were issued on dates when the market price of Apple stock was higher than actual market price on the actual grant dates. This improper backdating masked the virtually instant profits the option recipients obtained.On November 12, 2010, the settlement amount was amended to $16.5 million. In addition, Apple has agreed to implement certain corporate governance measures.
Caterpillar Inc	2017	NYSE	Ongoing	1471	According to the law firm press release, Caterpillar designs, manufactures, and markets construction, mining, and forestry machinery. Caterpillar distributes its products through a worldwide organization of dealers. The Complaint alleges that throughout the Class Period, Defendants made materially false and/or misleading statements, as well as failed to disclose material adverse facts about the Company's business, operations, and prospects. Specifically, Defendants made false and/or misleading statements and/or failed to disclose that: (i) Caterpillar unlawfully used foreign subsidiaries to avoid paying billions of dollars in US taxes; (ii) discovery of the foregoing conduct would subject the Company to heightened regulatory scrutiny and potential criminal sanctions; and (iii) as a result of the foregoing, Caterpillar's public statements were materially false and misleading at all relevant times.

Table B.2 continued from previous page

Firm	Year	Market	Case Status	Class Period Days	Summary
FleetCor Technologies Inc	2017	NYSE	Ongoing	365	The Complaint alleges that FleetCor falsely stated that the Company discloses its fees to customers clearly and that it focuses its business on helping employers control spending and save money. In truth, the Company owes its ostensible success to overcharging customers, disseminating misleading marketing materials, and engaging in predatory sales practices. In addition, FleetCor's contracts did not clearly disclose the Company's fees and FleetCor's improper business practices did not help customers control spending or save money.
General Motors	2014	NYSE	Settled	1209	The Complaint alleges that Defendants engaged in a scheme to hide from consumers and investors that GM's cars were plagued with a number of dangerous defects, which resulted in a multitude of adverse events including fatal car crashes and devastating injuries to GM's automobiles, drivers, passengers and others. On March 11, 2014, the Company sent a letter to the NHTSA detailing the alleged issues with GM's ignition switches and submitted a chronology of the Company's discovery of such problems and failure to properly advise consumers and regulators, leading to various injuries and deaths. In a span of less than seven weeks, the Company announced recalls more than 3.1 million vehicles. The Company also announced a $300 million charge in the first quarter of 2014 in order to deal with the recall campaigns.

Table B.2 continued from previous page

Firm	Year	Market	Case Status	Class Period Days	Summary
Pfizer Inc	2010	NYSE	Settled	1283	During the Class Period, defendants misled investors by failing to disclose that they were engaged in an ongoing course of conduct designed to illegally promote the sale of Pfizer drugs, including Bextra, Geodon, Zyvox and Lyrica. By such conduct, Pfizer caused hundreds of millions of dollars in false or fraudulent claims to be submitted to several federal healthcare programs, thus exposing the Company to untold legal liability. On January 26, 2009, Pfizer announced that it was paying $2.3 billion to resolve several ongoing investigations.
Walmart Inc	2012	NYSE	Settled	252	The complaint alleges defendants engaged in unlawful and unethical conduct. During the Class Period, defendants issued materially false and misleading statements regarding Walmart's practices with respect to unlawful or unethical conduct. Specifically, the Company failed to disclose that its executives had been involved in a multi-million-dollar bribery scheme at Walmart's Mexican subsidiary, Wal-Mart de Mexico. On September 24, 2018, the litigation was resolved. The Settling Parties' agreed to settle and release all claims that were asserted or could have been asserted in the Litigation in return for a cash payment of $160,000,000 to be paid by Walmart.

Notes: This table shows exemplary class action lawsuits that are included in our sample. These cases are intended to give an idea of the issues that are negotiated in class action lawsuits. The information was obtained from the SCAC and the case status was last updated on the day of data retrieval, February 6, 2019.

B.3 Distribution of Sample Firms Across Sectors

Table B.3: Distribution of Sample Firms Across Sectors

Sector	No. of Firms
Communication Services	17
Consumer Discretionary	276
Consumer Staples	59
Energy	87
Financials	354
Health Care	230
Industrials	274
Information Technology	250
Materials	69
Real Estate	143
Utilities	37

Notes: This table reports sector frequencies for our 1796 unique sample firms using the Global Industry Classification Standard.

B.4 Firm Percentages Facing Litigation

Table B.4: Firm Percentages Facing Litigation

	Whole sample	FPS sub-sample	Difference (Δ)
ClassPeriod	7.64%	9.69%	26.84%
CaseFiled	3.10%	4.27%	37.48%
ClassPeriodStart	2.76%	3.74%	35.29%

Notes: This table reports the percentages of firm-years from the whole sample and from the FPS sub-sample that faced a class period, case filing, and class period start. The third column reports the difference between the two groups.

B.5 Distribution of Class Action Lawsuits per Year

Table B.5: Distribution of Class Action Lawsuits per Year

	ClassPeriod		CaseFiled		ClassPeriodStart	
Year	absolute	in %	absolute	in %	absolute	in %
2003	26	13.20%	10	5.08%	11	5.58%
2004	25	9.06%	9	3.26%	7	2.54%
2005	30	9.93%	10	3.31%	8	2.65%
2006	28	9.06%	9	2.91%	12	3.88%
2007	35	10.67%	11	3.35%	18	5.49%
2008	36	8.18%	23	5.23%	11	2.50%
2009	31	6.08%	12	2.35%	14	2.75%
2010	32	5.94%	13	2.41%	13	2.41%
2011	32	5.86%	10	1.83%	12	2.20%
2012	43	7.69%	19	3.40%	22	3.94%
2013	45	7.84%	18	3.14%	19	3.31%
2014	39	6.52%	11	1.84%	11	1.84%
2015	76	7.12%	15	1.40%	31	2.90%
2016	135	8.16%	59	3.57%	48	2.90%
2017	121	6.79%	74	4.15%	39	2.19%

Notes: This table reports absolute numbers and percentages of case filings, class periods, and class period starts for the sample period. The percentages were calculated as the absolute number divided by the number of observations in the respective year.

B.6 Robustness: Logit – Estimation of Litigation Risk

Table B.6: Robustness Check – Estimation of Litigation Risk in Logit Regressions

	Class Period	Case Filed	Class Period Start
Independent Variables	(1)	(2)	(3)
ESG_P-score $_{t-1}$	−0.30	−0.43	−0.84**
	(0.52)	(0.50)	(0.47)
ESG_N-score $_{t-1}$	−0.83**	−1.22**	−1.05**
	(0.25)	(0.35)	(0.35)
Control Variables			
FPS	0.61**	0.67**	0.56**
	(0.17)	(0.16)	(0.16)
Ret_{t-1}	0.80	−1.55	5.83
	(2.21)	(2.75)	(3.73)
RetSkew_{t-1}	4.16	−8.18	8.10
	(6.84)	(10.65)	(10.21)
RetStdDev_{t-1}	56.67**	54.47**	34.85**
	(9.45)	(10.56)	(13.16)
SalesGrowth_{t-1}	1.63**	1.92**	1.39**
	(0.41)	(0.48)	(0.56)
Size_{t-1}	0.38**	0.35**	0.22**
	(0.07)	(0.06)	(0.06)
Turnover_{t-1}	−23.61**	−6.27	−19.38
	(9.85)	(12.21)	(15.57)
Intercept	−5.62**	−6.16**	−5.43**
	(0.68)	(0.73)	(0.78)
Year FE	Yes	Yes	Yes
ME ESG_N-score	−0.049	−0.027	−0.024
ME FPS	0.036	0.015	0.013
N	7671	7671	7671
F-statistic	6.930	7.579	3.482
p-value	0.000	0.000	0.000

Notes: This table shows the results from regressing ClassPeriod, CaseFiled, and ClassPeriodStart on ESG_N-score, ESG_P-score, and several control variables. In contrast to our main analysis, we used Logit panel regressions. Similar to our main analysis, we adjusted the standard errors that were used to calculate the test statistics and significance levels by a two-dimensional cluster to account for possible correlation of residuals across firm and time. Marginal effects (ME) for ESG_N-score and FPS are displayed at the bottom of the table. Standard errors in parentheses. * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

B.7 Robustness: Baseline Estimation of Litigation Risk

Table B.7: Baseline Estimation of Litigation Risk in Multiple Panel Regressions

	Class Period	Case Filed	Class Period Start
FPS	0.32***	0.32***	0.25***
	(0.08)	(0.07)	(0.07)
Size	0.20***	0.18***	0.10***
	(0.03)	(0.03)	(0.02)
SalesGrowth	0.72***	0.80***	0.55**
	(0.21)	(0.22)	(0.25)
Return	0.51	−0.64	2.37
	(1.06)	(1.23)	(1.57)
RetSkew	2.34	−3.00	3.80
	(3.42)	(4.71)	(4.43)
RetStdDev	30.86***	28.82***	15.56**
	(5.25)	(5.51)	(6.29)
Turnover	−13.11***	−5.01	−7.80
	(5.08)	(6.05)	(6.73)
Intercept	−3.38***	−3.63***	−3.17***
	(0.33)	(0.29)	(0.31)
McFadden Pseudo-R2 (%)	6.053	6.836	3.456
adj. McFadden Pseudo-R2 (%)	5.038	4.856	1.291
Year FE	Yes	Yes	Yes
ME FPS	0.093	0.045	0.031
N	7671	7671	7671
F-statistic	5.984	6.129	2.926
p-value	0.000	0.000	0.000

Notes: This table shows the results from regressing ClassPeriod, CaseFiled, and ClassPeriodStart on the variables used in the most cost-benefit-efficient model by Kim & Skinner (2012) (i.e., the "best" model). We used Probit panel regressions and adjusted the standard errors that were used to calculate the test statistics and significance levels by a two-dimensional cluster to account for possible correlation of residuals across firm and time (Cameron et al., 2011). Marginal effects (ME) FPS are displayed at the bottom of the table. * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

B.8 Detailed Cumulative Abnormal Returns

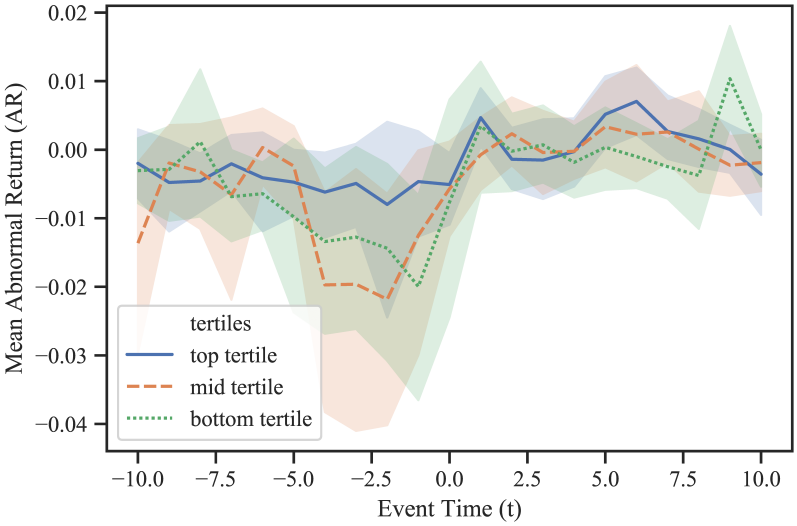
Table B.8: Cumulative Abnormal Returns

	Mean AR			Mean CAR		
	Top	Mid	Bottom	Top	Mid	Bottom
t_{-10}	−0.0024	−0.0136	−0.0031	−0.0024	−0.0136	−0.0031
	(−0.84)	(−1.93)	(−1.27)	(−0.84)	(−1.93)	(−1.27)
t_{-9}	−0.0041	−0.0019	−0.0029	−0.0065	−0.0156*	−0.0059
	(−1.18)	(−0.6)	(−0.81)	(−1.42)	(−2.01)	(−1.47)
t_{-8}	−0.0048*	−0.0032	0.0012	−0.0112*	−0.0188*	−0.0048
	(−2.23)	(−0.85)	(0.22)	(−2.16)	(−2.26)	(−0.73)
t_{-7}	−0.0022	−0.0066	−0.0068	−0.0135**	−0.0254*	−0.0116
	(−1.07)	(−0.95)	(−1.92)	(−2.64)	(−2.13)	(−1.64)
t_{-6}	−0.0042	0.0003	−0.0064*	−0.0177*	−0.025	−0.018*
	(−1.13)	(0.12)	(−2.56)	(−2.57)	(−1.89)	(−2.33)
t_{-5}	−0.006*	−0.0024	−0.0098	−0.0237**	−0.0274*	−0.0278**
	(−2.26)	(−0.66)	(−1.58)	(−2.87)	(−2.06)	(−2.62)
t_{-4}	−0.0063*	−0.0197*	−0.0134*	−0.03**	−0.0472**	−0.0412**
	(−2.0)	(−2.4)	(−2.11)	(−3.4)	(−2.97)	(−3.26)
t_{-3}	−0.0047	−0.0196	−0.0128	−0.0347**	−0.0668**	−0.054**
	(−1.4)	(−1.91)	(−1.94)	(−3.53)	(−3.65)	(−3.99)
t_{-2}	−0.008	−0.0218*	−0.0144*	−0.0426**	−0.0886**	−0.0684**
	(−1.07)	(−2.56)	(−2.02)	(−3.06)	(−4.18)	(−4.38)
t_{-1}	−0.0043	−0.0125	−0.02**	−0.047**	−0.1011**	−0.0883**
	(−1.1)	(−1.67)	(−2.7)	(−3.65)	(−4.76)	(−5.25)
t_0	−0.0057*	−0.0057	−0.0076	−0.0527**	−0.1068**	−0.096**
	(−2.1)	(−1.6)	(−0.89)	(−3.87)	(−5.0)	(−5.2)
t_1	0.0046*	−0.0007	0.0034	−0.0481**	−0.1076**	−0.0925**
	(2.21)	(−0.26)	(0.71)	(−3.55)	(−4.86)	(−4.85)
t_2	−0.0011	0.0023	−0.0002	−0.0492**	−0.1052**	−0.0928**
	(−0.43)	(0.88)	(−0.07)	(−3.52)	(−4.89)	(−4.98)
t_3	−0.0009	−0.0004	0.0007	−0.05**	−0.1057**	−0.0921**
	(−0.28)	(−0.13)	(0.23)	(−3.34)	(−4.92)	(−4.91)
t_4	0.0	−0.0002	−0.0018	−0.05**	−0.1059**	−0.0939**
	(0.01)	(−0.1)	(−0.69)	(−3.34)	(−4.87)	(−4.99)
t_5	0.0054	0.0033	0.0004	−0.0446**	−0.1025**	−0.0935**
	(1.86)	(1.03)	(0.11)	(−3.3)	(−4.77)	(−4.89)
t_6	0.0069*	0.0022	−0.001	−0.0377**	−0.1003**	−0.0945**
	(2.48)	(0.52)	(−0.4)	(−3.03)	(−4.92)	(−4.89)
t_7	0.0022	0.0026	−0.0025	−0.0356**	−0.0977**	−0.097**
	(0.85)	(1.17)	(−1.12)	(−2.76)	(−4.76)	(−4.93)
t_8	0.0014	0.0001	−0.0038	−0.0342*	−0.0976**	−0.1008**
	(0.7)	(0.02)	(−0.95)	(−2.54)	(−4.68)	(−4.8)
t_9	−0.0006	−0.0023	0.0103**	−0.0347**	−0.0999**	−0.0904**
	(−0.31)	(−1.0)	(2.62)	(−2.58)	(−4.7)	(−4.4)
t_{10}	−0.0043	−0.0019	0.0	−0.0391*	−0.1018**	−0.0905**
	(−1.53)	(−0.86)	(−0.02)	(−2.57)	(−4.73)	(−4.37)

Notes: This table displays mean abnormal returns (Mean AR) and mean cumulative abnormal returns (Mean CAR) for a total event window of 21 days $[-10,+10]$. We split the sample into tertiles (Top, Mid, and Bottom) by the firms' event-year ESG_P-score. In parentheses, we report standardized cross-sectional t-statistics (Boehmer et al., 1991) for AR and cross-sectional t-statistics for CAR. * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

B.9 Mean Abnormal Returns – Tertiles

Figure B.1: Mean Abnormal Returns



Notes: This graph shows mean abnormal returns for the firms from the top, mid, and bottom tertiles. The shaded areas indicate 95% confidence intervals.

B.10 Robustness: Alternative Models for Estimating Abnormal Returns

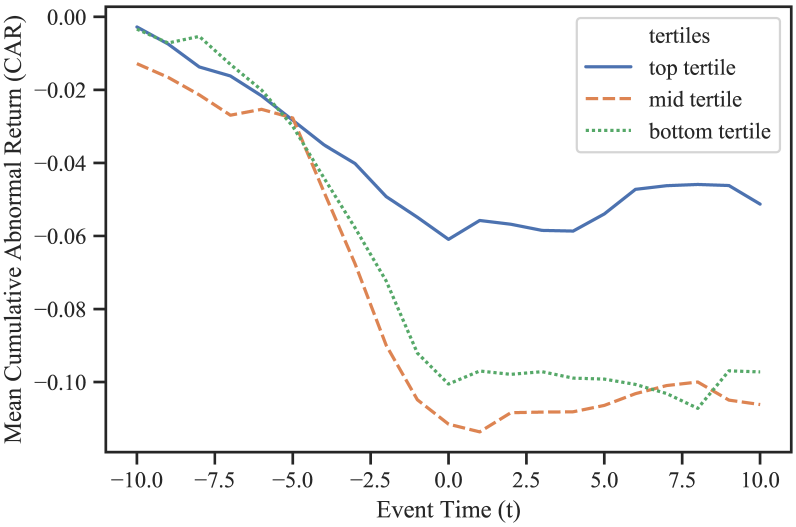
Table B.9: Alternative Models for the Event Study Estimation of Abnormal Returns

		ESG Tertiles		
		Top	Mid	Bottom
Fama-French Model	Mean $CAR_{t=10}$	-0.0391^{**}	-0.1018^{**}	-0.0905^{**}
	Stand. t-statistic	-3.33	-5.11	-4.87
	Patell Z	-5.06	-10.53	-7.83
Carhart Model	Mean $CAR_{t=10}$	-0.0367^{**}	-0.0988^{**}	-0.0836^{**}
	Stand. t-statistic	-3.18	-5.04	-4.37
	Patell Z	-5.54	-10.08	-9.23
CRSP Market Model	Mean $CAR_{t=10}$	-0.0455^{**}	-0.0972^{**}	-0.0879^{**}
	Stand. t-statistic	-3.66	-5.10	-4.90
	Patell Z	-5.37	-10.07	-7.44

Notes: This table shows the results for mean $CAR_{t=10}$ from our base Fama and French model together with the alternative Carhart four-factor model and the CRSP value-weighted market model as robustness checks. We used standardized t-statistics (Boehmer et al., 1991). * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

B.11 Robustness: Alternative Estimation Window

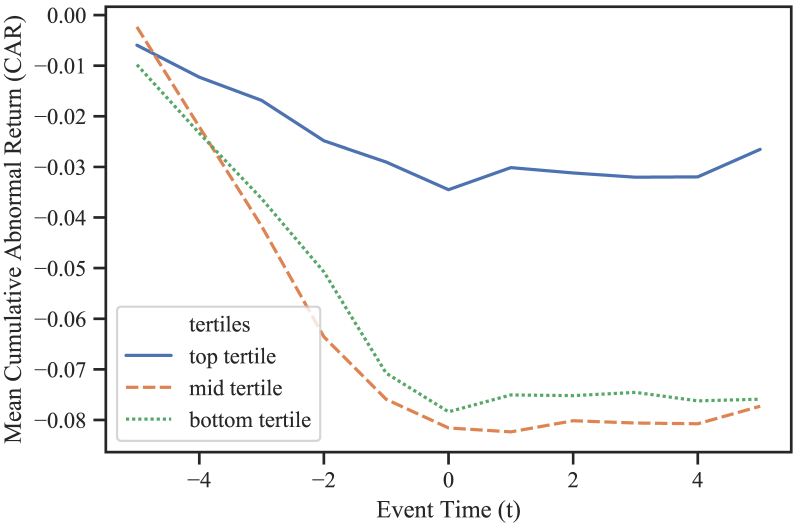
Figure B.2: Alternative Estimation Window



Notes: In this graph, we increased the estimation window to 500 days and the gap between the estimation and event window to 250 days.

B.12 Robustness: Alternative Event Window

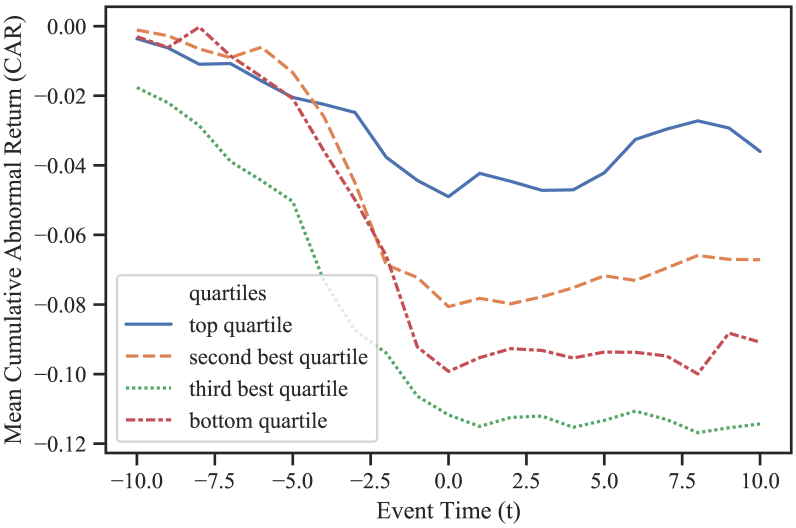
Figure B.3: Alternative Event Window



Notes: In this graph, we shortened the event window to 11 days $[-5, +5]$.

B.13 Robustness: Alternative Cut-Offs

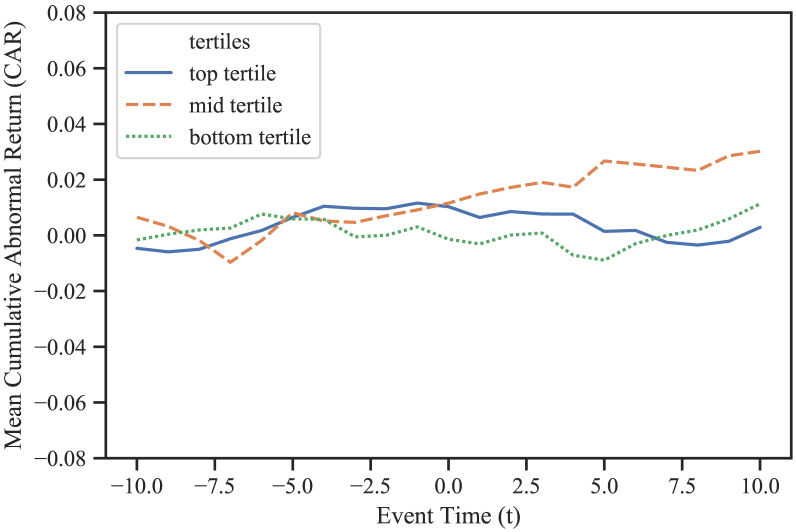
Figure B.4: Alternative Cut-Offs



Notes: In this graph, we used alternative cut-offs and thus assigned the firms in our sample into quartiles based on their event-year ESG_P-score.

B.14 Robustness: Placebo Event Study

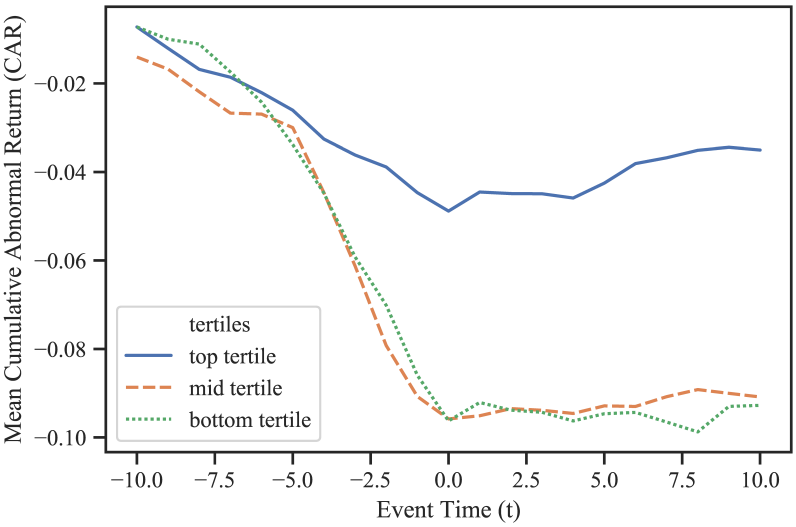
Figure B.5: Placebo Event Study



Notes: In this graph, we shifted the real event date by one year into the future to create an artificial event date on which no class action lawsuit was filed.

B.15 Robustness: Winsorized CAR

Figure B.6: Winsorized CAR



Notes: In this graph, we winsorized CAR that were used to calculate the mean CAR to control for a possible outlier effect on our results.

B.16 Mean CAR per Industry

Table B.10: Mean $CAR_{t=10}$ per Industry

Sector	Mean $CAR_{t=10}$
Communication Services	−0.249
Consumer Discretionary	−0.032
Consumer Staples	0.016
Energy	−0.127
Financials	−0.063
Health Care	−0.063
Industrials	−0.069
Information Technology	−0.024
Materials	−0.118
Real Estate	−0.115
Utilities	−0.120

Notes: This table reports the mean $CAR_{t=10}$ of treated firms clustered by industry.

B.17 Robustness: Monthly Fama-French 5-Factor Alphas

Table B.11: Robustness: Base Trading Strategies – Monthly Fama-French 5-Factor Alphas

	Single-Sorted				Double-Sorted
	Low Lit	High Lit	High ESG_p	Low ESG_p	Low Lit-High ESG_p
Panel Equally-Weighted					
Alpha	0.0067**	−0.0130**	0.0004	0.0035*	0.0162**
Adj. R2	0.8356	0.7046	0.9307	0.8908	0.5866
Turnover	0.755	0.580	0.379	0.600	0.977
Panel Value-Weighted					
Alpha	0.0092**	0.0015	0.0025**	0.0050*	0.0142**
Adj. R2	0.8076	0.9417	0.9585	0.7705	0.5144
Turnover	0.752	0.242	0.259	0.541	0.980

Notes: This table reports monthly Fama-French 5-factor alphas of the single-sorted portfolios: "Low Lit," "High Lit," "High ESG_p," and "Low ESG_p." Moreover, the last column of the table contains the monthly Fama-French 5-factor alpha of the double-sorted portfolio, "Low Lit-High ESG_p." We used 10% cutoffs for the litigation probability as well as ESG_P-score to build the portfolios. The backtesting time frame was 2014 to 2017 (N=48). Moreover, we report annual turnover in per cent. * Statistically significant at the 5% level; ** Statistically significant at the 1% level.

C Appendix: Paper 3

C.1 Specification of Optimization Problems

This section outlines the optimization problems being solved by the algorithms. It also states relevant references, which can be referred to for further details on the optimization problems and algorithms.

Linear Regression

Our linear model had the following simple form:

$$\hat{y} = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \theta_3 \cdot x_3 + ... + \theta_n \cdot x_n \tag{C.1.1}$$

where $\hat{y}$ is the predicted value; x_i is the ith predictor value; θ_m is the mth model parameter; and n is the number of predictors. We thus ended up with the following simple optimization problem:

$$\min MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{C.1.2}$$

Elastic Net

We followed Zou and Hastie (2005) and Friedman et al. (2010) by specifying the Elastic Net optimization problem as follows:

$$\min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} R_\lambda(\beta_0,\beta) = \min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} \left[\frac{1}{2N}\sum_{i=1}^{N}(y_i-\beta_0-x_i^\top\beta)^2+\lambda P_\alpha(\beta)\right],$$

where (C.1.3)

$$P_\alpha(\beta) = \sum_{j=1}^{p}\left[\frac{1}{2}(1-\alpha)\beta_j^2+\alpha\mid\beta_j\mid\right]$$

where P_α is the elastic net penalty (i.e., a combination of the lasso and ridge penalty). An α of 1 is equivalent to ridge regression, while an α of 0 is equivalent to lasso regression. Thus, alphas between 1 and 0 lead to an elastic net penalty. In the algorithm default settings of *sklearn*, α is 0.5.[80]

Ridge

We specified the optimization problem in the form of the cost function that we are trying to minimize as follows (Hoerl, 1962; Hoerl and Kennard, 1970):

$$\min_{(w)\in\mathbb{R}^{p+1}} cost(w) = \frac{1}{2*n}\sum_{i=1}^{i=n}(y_i-\hat{y}_i)^2+\alpha\sum_{j=1}^{j=D}\mid w_j\mid \tag{C.1.4}$$

where $\alpha\sum_{j=1}^{j=D}\mid w_j\mid$ corresponds to the $\ell 2$ penalty, which applies the regularization. α is a hyper-parameter that has to be determined by the researcher (an α of 0 is equal to linear regression). A higher α leads to stronger regularization. In our case, the default setting for α in *sklearn* is $\alpha = 1$.

Support Vector Regression

We followed Chang and Lin (2011) and Vapnik (1998) and specified the SVR optimization problem as follows:

[80] Please note that in *sklearn*, l1_ratio corresponds to alpha rather than to the hyper-parameter α. *sklearn* uses a coordinate descent solver for the optimization problem. For more information on the coordinate descent solver, please see Friedman et al. (2010) and Tseng (2001)

$$\begin{aligned} \min_{(w,b,\xi,\xi^*)} \quad & \frac{1}{2} w^\top w + C \sum_{i=1}^{n} \xi_i + C \sum_{i=1}^{n} \xi_i^*, \\ \text{subject to} \quad & w^\top \phi(x_i) + b - y_i \leq \varepsilon + \xi_i, \\ & y_i - w^\top \phi(x_i) - b \leq \varepsilon + \xi_i^*, \\ & \xi_i, \xi_i^* \geq 0, i = 1, ..., n \end{aligned} \tag{C.1.5}$$

for an array of training observations, $(x_1, y_1), ..., (x_n, y_n)$, where $x_i \in \mathbb{R}^n$ is a vector of predictors and $y_i \in \mathbb{R}^n$ is the output value, under given parameters $C > 0$ and $\varepsilon > 0$. The two kernel functions implemented in this study take the following form (Chang and Lin, 2011; Barboza et al., 2017):

$$K(x_i, x_j) = < x_i, x_j > \tag{C.1.6}$$

for the linear kernel and

$$K(x_i, x_j) = \exp(-\gamma \parallel x_i - x_j \parallel^2) \tag{C.1.7}$$

for the RBF kernel. While the linear kernel does not work very well on non-linear data, the RBF kernel is capable of capturing non-linear relationships in the data. Therefore, the prediction accuracy of RbfSVR is usually superior to that of LinearSVR. Computationally, however, RbfSVR is more expensive and thus comes at the cost of increased training time.

Decision Tree Regression

The logic behind the used CART algorithm is quite intuitive. More specifically, the algorithm splits the training set into two subsets based on a single predictor p (e.g., liquidity) and a certain threshold t_p (e.g., > 4.8), which it determines by searching for the pair of p and t_p that produces the purest subsets (Géron, 2017) (i.e., minimizes the mean squared error, MSE). The cost function of the CART regression algorithm is as follows (Géron, 2017):

$$\begin{aligned} \min \quad & J(p,t_p) = \frac{m_{left}}{m} MSE_{left} + \frac{m_{right}}{m} MSE_{right}, \\ & \text{where} \\ & MSE_{node} = \sum_{i \in node} (\hat{y}_{node} - y^{(i)})^2, \\ & \hat{y}_{node} = \frac{1}{m_{node}} \sum_{i \in node} y^{(i)} \end{aligned} \tag{C.1.8}$$

The CART algorithm then recursively splits the subsets, sub-subsets, sub-sub-subsets, etc., until the cost function is minimized or reaches a hyper-parameter threshold (Géron, 2017). DTR belongs to that type of algorithm that is most prone to over-fitting the training data. However, as explained, we used *sklearn's* default package settings and thus refrained from hyper-parameter tuning. Besides, our next algorithm, Random Forest regression, presents an alternative tool to avoid over-fitting.

Random Forest Regression

Using bagging, our RFR algorithm solved the following regression prediction task (Breiman, 1996, 2001):

$$\hat{r} = \frac{1}{T} \sum_{t=1}^{T} r_t(x') \tag{C.1.9}$$

where T equals the number of bootstrapped samples; r_t are the trained regression trees on the T-times bootstrapped samples, X_t and Y_t; and x' denotes unseen input data. Put simply, RFR averages the predictions of T trained regression trees.

AdaBoost Regression

Our AdaBoostR algorithm followed a modified version of Freund and Schapire's (1997) *Adaboost.R*, introduced by Drucker (1997):

Before running the algorithm, a distribution of training instance weights $D_i = \frac{w_i}{\sum_{i=1}^{N} w_i}$

with $w_i = 1$ and $i = 1,...,N_1$ is instantiated. The following steps are then repeated as long as the average loss $\overline{L}$ is less then 0.5.

1. Draw N_1 bootstrapped samples (with replacement) from the training set. Build a new regression machine "a" and run it on the training set to obtain a prediction $\hat{y}_i^{(D)}$.
2. Calculate a linear loss $L_i = \frac{|\hat{y}_i^{(D)}(x_i) - y_i|}{\sup|\hat{y}_i^{(D)}(x_i) - y_i|}$ for each sample $i = 1,...,N_1$ and with $L \in [0,1]$
3. Average loss and the confidence measure are calculated as $\overline{L} = \sum_{i=1}^{N_1} L_i D_i$ and $\beta = \frac{\overline{L}}{1-\overline{L}}$, respectively. Low β denotes a highly confident prediction.
4. Next, weights are updated as follows: $w_i \rightarrow w_i \beta exp(1 - L_i)$. Smaller losses lead to larger decreases in weights and thus to a lower probability that this instance will again be picked as a member of the training set in the next ensemble machine.

Each of the A regression machines makes a prediction $\hat{y}_a$ for $a = 1,...,A$ and input x_i. The cumulative prediction $\hat{y}_f$ is obtained as follows:

$$\hat{y}_f = \inf\left\{ y \in Y : \sum_{a:\hat{y}_a \leq y} \log(1/\beta_a) \geq \frac{1}{2} \sum_a \log(1/\beta_a) \right\} \tag{C.1.10}$$

Gradient Boosting Regression

We followed Friedman (2001) and used GBR with a least-squares loss function to solve the following optimization problem:[81]

$$\begin{aligned} F^*(x) &= \underset{F}{\operatorname{argmin}} E_x[E_y(L(y,F(x))) \mid x] \\ &\text{where} \\ L(y,F(x)) &= \frac{1}{2}(y - F)^2 \end{aligned} \tag{C.1.11}$$

[81] Due to the similarity to AdaBoostR, we refrained from describing every single step of the algorithm. Please refer to Friedman (2001) for more information.

K-nearest Neighbors Regression

Our regression algorithm uses the Minkowski distance measure to determine the k (in our case, 5) closest neighbors of a new data point. The Minkowski distance measure for a set of data points $(x_1, y_1), ..., (x_n, y_n) \in \mathbb{R}^n$ is calculated as follows:

$$\left[\sum_{i=1}^{k} (\mid x_i - y_i \mid)^q\right]^{1/q} \quad \text{with} \quad q = 2 \quad \text{(“Euclidean distance”)} \tag{C.1.12}$$

Multi-layer Perceptron Regression

The Adam weights update rule can be described as follows (Kingma and Ba, 2015; Ruder, 2017):[82]

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{\upsilon}_t} + \varepsilon} \hat{m}_t \quad \text{where} \quad \hat{m}_t = \frac{\beta_1 m_{t-1} + (1-\beta_1) g_t}{1-\beta_1^t} \quad \hat{\upsilon}_t = \frac{\beta_2 \upsilon_{t-1} + (1-\beta_2) g_t^2}{1-\beta_2^t} \tag{C.1.13}$$

$\hat{\upsilon}_t$ and $\hat{m}_t$ are the bias-corrected first and second moment estimates (Ruder, 2017), i.e., the mean and uncentered variance, respectively.

[82] Please see Kingma and Ba (2015) for a detailed pseudo-code of Adam.

C.2 Variable Definitions

Table C.12: Variable Definitions

Variable	Type	Description
Credit Risk	$[1;22]$	measures credit risk by S&P's long-term issuer credit ratings. The original ratings, which ranged from AAA to D, were converted into a numerical and continuous variable (1 for AAA, 2 for AA, 3 for A, 4 for BBB, etc.). Thus, the lowest possible rating (i.e., D) resulted in a value of 22.
Current RRI	$[0;100]$	measures current exposure to CSI, based on a proprietary RepRisk algorithm. This paper uses quarterly observations. Higher (lower) values indicate higher (lower) CSI exposure compared to the firm's peers.
Peak RRI	$[0;100]$	measures the highest level of CSI exposure over the last two years, based on a proprietary RepRisk algorithm. Higher (lower) values indicate higher (lower) CSI exposure compared to the firm's peers.
ESG Issues	$[0;\infty]$	is a vector containing all 28 ESG issues from RepRisk that ultimately constitute the RRI. This vector is multiplied with the news count for every issue and firm-quarter observation. For instance, if a firm experienced a forced labor incident and the news count was 8 in that respective firm-quarter, the ESG issue "Forced labor" took the value of $1 \cdot 8 = 8$. Summary statistics can be found in Table 15. RepRisk provides more information, please see: https://www.reprisk.com/content/static/reprisk-esg-issues-definitions.pdf
LIQU	$[-\infty;1]$	measures liquidity. Calculated as working capital divided by total assets.
PROF	$[-\infty;\infty]$	measures profitability. Calculated as retained earnings divided by total assets.
OPEF	$[-\infty;\infty]$	measures operating efficiency. Calculated as earnings before interest and taxes (EBIT) divided by total assets.
ME	$[-\infty;\infty]$	accounts for a market effect on credit risk. Calculated by taking the natural logarithm of market value divided by long-term debt.
AT	$[0;\infty]$	measures asset turnover. Calculated as sales divided by total assets.
ASSETS	$[-1;\infty]$	measures asset growth. Calculated as the percentage change in total assets between the current and last quarter.
SALES	$[-1;\infty]$	measures sales growth. Calculated as the percentage change in sales between the current and last quarter.
P/B	$[-1;\infty]$	measures the change in the price-to-book-value ratio. Calculated as the percentage change in the price-to-book-value ratio between the current and last quarter.
LEV	$[-\infty;\infty]$	measures leverage. Calculated as long-term debt divided by stockholders equity.
AGE	$[0;\infty]$	measures firm age. Calculated as the current year minus the year in which the firm appeared in the Compustat/CRSP database for the first time.
SIZE	$[0;\infty]$	measures firm size. Calculated as the natural logarithm of total assets.
INTCOV	$[-\infty;\infty]$	measures interest coverage. Calculated by taking the natural logarithm of operating income before extraordinary items divided by interest expense.
CAPINT	$[0;1]$	measures capital intensity. Calculated as net property, plant, and equipment (PPE) divided by total assets.

C.3 Algorithm Default Hyper-Parameters

Table C.13: Algorithm Default Hyper-Parameters

Algorithm	Library	Description
Linear	sklearn.linear_model	*class* LinearRegression (fit_intercept=True, normalize=False, copy_X=True, n_jobs=None)
ElasticNet	sklearn.linear_model	*class* ElasticNet (alpha=1.0, l1_ratio=0.5, fit_intercept=True, normalize=False, precompute=False, max_iter=1000, copy_X=True, tol=0.0001, warm_start=False, positive=False, random_state=target_seed_global, selection='cyclic')
Ridge	sklearn.linear_model	*class* Ridge (alpha=1.0, fit_intercept=True, normalize=False, copy_X=True, max_iter=None, tol=0.001, solver='auto', random_state=target_seed_global)
OrdinalRidge	mord	*class* OrdinalRidge (alpha=1.0, fit_intercept=True, normalize=False, copy_X=True, max_iter=None, tol=0.001, solver='auto')
LinearSVR	sklearn.svm	*class* SVR (kernel='linear', degree=3, gamma='scale', coef0=0.0, tol=0.001, C=1.0, epsilon=0.1, shrinking=True, cache_size=200, verbose=False, max_iter=-1)
RbfSVR	sklearn.svm	*class* SVR (kernel='rbf', degree=3, gamma='auto', coef0=0.0, tol=0.001, C=1.0, epsilon=0.1, shrinking=True, cache_size=200, verbose=False, max_iter=-1)
DTR	sklearn.tree	*class* DecisionTreeRegressor (criterion='mse', splitter='best', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=None, random_state=target_seed_global, max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, presort='deprecated', ccp_alpha=0.0)
RFR	sklearn.ensemble	*class* RandomForestRegressor (n_estimators=10, criterion='mse', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, bootstrap=True, oob_score=False, n_jobs=None, random_state=target_seed_global, verbose=0, warm_start=False, ccp_alpha=0.0, max_samples=None)
AdaBoostR	sklearn.ensemble	*class* AdaBoostRegressor (base_estimator=None, n_estimators=50, learning_rate=1.0, loss='linear', random_state=target_seed_global)
GBR	sklearn.ensemble	*class* GradientBoostingRegressor (loss='ls', learning_rate=0.1, n_estimators=100, subsample=1.0, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_depth=3, init=None, random_state=target_seed_global, max_features=None, alpha=0.9, verbose=0, max_leaf_nodes=None, warm_start=False)
MLPR	sklearn.neural_network	*class* MLPRegressor (hidden_layer_sizes=(100,), activation='relu', solver='adam', alpha=0.0001, batch_size='auto', learning_rate='constant', learning_rate_init=0.001, power_t=0.5, max_iter=1000000, shuffle=True, random_state=target_seed_global, tol=0.0001, verbose=False, warm_start=False, momentum=0.9, nesterovs_momentum=True, early_stopping=False, validation_fraction=0.1, beta_1=0.9, beta_2=0.999, epsilon=1e-08, n_iter_no_change=10, max_fun=15000)
KNNR	sklearn.neighbors	*class* KNeighborsRegressor (n_neighbors=5, weights='uniform', algorithm='auto', leaf_size=30, p=2, metric='minkowski', metric_params=None, n_jobs=None, **kwargs)

C.4 Robustness: US Data

Table C.14: Robustness Check – US Data

Panel A: Baseline – Without CSI Exposure

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.6868	0.6865	1.7952	Linear	0.6942	0.6929	1.7607
ElasticNet	0.5451	0.5446	2.1635	ElasticNet	0.5469	0.5449	2.1433
Ridge	0.6868	0.6864	1.7953	Ridge	0.6940	0.6926	1.7614
OrdinalRidge	0.6797	0.6793	1.8155	OrdinalRidge	0.6865	0.6852	1.7826
LinearSVR	0.6802	0.6799	1.8139	LinearSVR	0.6867	0.6854	1.7821
RbfSVR	0.8472	0.8471	1.2538	RbfSVR	0.8398	0.8391	1.2742
DTR	0.9988	0.9988	0.1133	DTR	0.8330	0.8323	1.3012
RFR	0.9847	0.9847	0.3966	RFR	0.9247	0.9244	0.8738
AdaBoostR	0.6639	0.6635	1.8598	AdaBoostR	0.6582	0.6567	1.8615
GBR	0.8211	0.8209	1.3567	GBR	0.8127	0.8119	1.3780
KNNR	0.9317	0.9316	0.8384	KNNR	0.8904	0.8899	1.0540
MLPR	0.8670	0.8668	1.1699	MLPR	0.8683	0.8677	1.1557

Panel B: Current RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.6869	0.6865	1.7950	Linear	0.6942	0.6928	1.7607
ElasticNet	0.5395	0.5390	2.1768	ElasticNet	0.5400	0.5379	2.1594
Ridge	0.6868	0.6864	1.7952	Ridge	0.6939	0.6926	1.7614
OrdinalRidge	0.6798	0.6794	1.8152	OrdinalRidge	0.6870	0.6856	1.7813
LinearSVR	0.6802	0.6799	1.8140	LinearSVR	0.6865	0.6851	1.7826
RbfSVR	0.8509	0.8508	1.2385	RbfSVR	0.8417	0.8410	1.2669
DTR	0.9998	0.9998	0.0419	DTR	0.8402	0.8395	1.2727
RFR	0.9853	0.9853	0.3885	RFR	0.9233	0.9229	0.8819
AdaBoostR	0.6675	0.6671	1.8497	AdaBoostR	0.6624	0.6609	1.8500
GBR	0.8206	0.8204	1.3586	GBR	0.8121	0.8112	1.3803
KNNR	0.8428	0.8427	1.2717	KNNR	0.7547	0.7536	1.5770
MLPR	0.8954	0.8953	1.0373	MLPR	0.8747	0.8741	1.1270

Table C.14 continued from the previous page

Panel C: Peak RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.6868	0.6865	1.7952	Linear	0.6942	0.6928	1.7607
ElasticNet	0.5398	0.5393	2.1760	ElasticNet	0.5404	0.5383	2.1585
Ridge	0.6868	0.6864	1.7953	Ridge	0.6940	0.6926	1.7613
OrdinalRidge	0.6795	0.6792	1.8159	OrdinalRidge	0.6865	0.6851	1.7826
LinearSVR	0.6803	0.6799	1.8138	LinearSVR	0.6867	0.6853	1.7821
RbfSVR	0.8519	0.8518	1.2344	RbfSVR	0.8432	0.8425	1.2606
DTR	0.9998	0.9998	0.0409	DTR	0.8333	0.8325	1.3001
RFR	0.9856	0.9855	0.3855	RFR	0.9256	0.9253	0.8682
AdaBoostR	0.6640	0.6636	1.8595	AdaBoostR	0.6529	0.6513	1.8758
GBR	0.8214	0.8212	1.3555	GBR	0.8127	0.8119	1.3778
KNNR	0.9011	0.9010	1.0089	KNNR	0.8277	0.8269	1.3217
MLPR	0.9043	0.9042	1.0179	MLPR	0.8849	0.8844	1.0800

Panel D: ESG Issues

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.6896	0.6889	1.7871	Linear	0.6955	0.6925	1.7570
ElasticNet	0.5451	0.5440	2.1635	ElasticNet	0.5469	0.5424	2.1433
Ridge	0.6896	0.6888	1.7873	Ridge	0.6952	0.6923	1.7577
OrdinalRidge	0.6822	0.6814	1.8085	OrdinalRidge	0.6890	0.6860	1.7755
LinearSVR	0.6827	0.6819	1.8070	LinearSVR	0.6880	0.6849	1.7785
RbfSVR	0.7910	0.7904	1.4666	RbfSVR	0.7697	0.7675	1.5278
DTR	0.9989	0.9989	0.1058	DTR	0.8342	0.8326	1.2963
RFR	0.9845	0.9845	0.3989	RFR	0.9241	0.9234	0.8770
AdaBoostR	0.6759	0.6751	1.8263	AdaBoostR	0.6705	0.6673	1.8277
GBR	0.8216	0.8211	1.3550	GBR	0.8130	0.8111	1.3770
KNNR	0.9252	0.9250	0.8773	KNNR	0.8792	0.8780	1.1068
MLPR	0.8986	0.8983	1.0216	MLPR	0.8259	0.8242	1.3284

Notes: This table displays the robustness results using US data. The table reports the three most common measures of prediction accuracy for all regression models: R^2, adj. R^2, and root mean squared error (RMSE). In Panel A, we used the baseline specification and thus no CSI predictor. In Panel B, we used Current RRI as a predictor to capture CSI exposure. In Panel C, we used Peak RRI as a predictor to capture CSI exposure. In Panel D, we used the vector of ESG issues as a predictor to capture CSI exposure. The vector of additional accounting- and market-related variables and the vector of sector dummies were included in all models. For more information on algorithm specifications, please see Table C.13.

C.5 Additional Analysis: ESG Issues with High Reach and High Severity

Table C.15: Additional Analysis – ESG Issues with High Reach and High Severity

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5606	0.5599	2.1173	Linear	0.5641	0.5613	2.1025
ElasticNet	0.4416	0.4407	2.3868	ElasticNet	0.4508	0.4473	2.3600
Ridge	0.5606	0.5599	2.1173	Ridge	0.5640	0.5612	2.1027
OrdinalRidge	0.5531	0.5524	2.1353	OrdinalRidge	0.5563	0.5534	2.1212
LinearSVR	0.5533	0.5526	2.1349	LinearSVR	0.5547	0.5518	2.1250
RbfSVR	0.7355	0.7350	1.6428	RbfSVR	0.7417	0.7401	1.6183
DTR	0.9990	0.9990	0.1031	DTR	0.8060	0.8048	1.4025
RFR	0.9787	0.9787	0.4663	RFR	0.8979	0.8972	1.0176
AdaBoostR	0.5831	0.5824	2.0625	AdaBoostR	0.5720	0.5692	2.0834
GBR	0.7356	0.7351	1.6425	GBR	0.7347	0.7330	1.6403
KNNR	0.9091	0.9090	0.9628	KNNR	0.8510	0.8501	1.2290
MLPR	0.7810	0.7806	1.4949	MLPR	0.7818	0.7804	1.4874

Notes: This table reports the three most common measures of prediction accuracy for all regression models: R^2, adj. R^2, and root mean squared error (RMSE). We used interaction terms between our vector of ESG issues, a high reach dummy variable, and a high severity dummy variable as a predictor to capture CSI exposure. The vector of additional accounting- and market-related variables and the vector of sector dummies were included in all models. For more information on algorithm specifications, please see Table C.13.

C.6 Robustness: Increased Test Sample Size

We increased the size of our test sample from 20% to 30% of the total data to mitigate concerns about a too-small test sample driving our results. However, increasing the test sample size did not qualitatively change our results (cf., Table C.16).

Table C.16: Robustness Check – Increased Test Size

Panel A: Baseline – Without CSI Exposure

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5605	0.5601	2.1148	Linear	0.5621	0.5613	2.1160
ElasticNet	0.4431	0.4426	2.3806	ElasticNet	0.4429	0.4418	2.3868
Ridge	0.5605	0.5601	2.1148	Ridge	0.5621	0.5613	2.1160
OrdinalRidge	0.5531	0.5528	2.1325	OrdinalRidge	0.5547	0.5539	2.1337
LinearSVR	0.5529	0.5526	2.1329	LinearSVR	0.5546	0.5537	2.1341
RbfSVR	0.7636	0.7634	1.5509	RbfSVR	0.7579	0.7574	1.5734
DTR	0.9993	0.9993	0.0852	DTR	0.7801	0.7797	1.4994
RFR	0.9773	0.9773	0.4806	RFR	0.8881	0.8879	1.0695
AdaBoostR	0.5869	0.5865	2.0503	AdaBoostR	0.5778	0.5769	2.0779
GBR	0.7363	0.7360	1.6383	GBR	0.7299	0.7294	1.6619
KNNR	0.9026	0.9025	0.9957	KNNR	0.8377	0.8374	1.2883
MLPR	0.8077	0.8076	1.3987	MLPR	0.8075	0.8071	1.4031

Panel B: Current RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5717	0.5713	2.0876	Linear	0.5740	0.5731	2.0871
ElasticNet	0.4515	0.4510	2.3626	ElasticNet	0.4511	0.4500	2.3690
Ridge	0.5717	0.5713	2.0876	Ridge	0.5740	0.5731	2.0871
OrdinalRidge	0.5649	0.5646	2.1041	OrdinalRidge	0.5658	0.5650	2.1070
LinearSVR	0.5653	0.5649	2.1032	LinearSVR	0.5682	0.5674	2.1011
RbfSVR	0.7710	0.7708	1.5265	RbfSVR	0.7654	0.7650	1.5487
DTR	0.9999	0.9999	0.0337	DTR	0.7786	0.7781	1.5048
RFR	0.9775	0.9775	0.4788	RFR	0.8858	0.8856	1.0806
AdaBoostR	0.6025	0.6021	2.0113	AdaBoostR	0.5965	0.5957	2.0313
GBR	0.7379	0.7377	1.6331	GBR	0.7316	0.7310	1.6567
KNNR	0.8048	0.8046	1.4094	KNNR	0.6862	0.6855	1.7914
MLPR	0.8037	0.8035	1.4134	MLPR	0.8051	0.8048	1.4115

Table C.16 continued from the previous page

Panel C: Peak RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5697	0.5693	2.0925	Linear	0.5720	0.5711	2.0921
ElasticNet	0.4499	0.4494	2.3660	ElasticNet	0.4497	0.4486	2.3720
Ridge	0.5697	0.5693	2.0925	Ridge	0.5719	0.5711	2.0921
OrdinalRidge	0.5612	0.5608	2.1132	OrdinalRidge	0.5643	0.5634	2.1107
LinearSVR	0.5631	0.5628	2.1084	LinearSVR	0.5661	0.5652	2.1063
RbfSVR	0.7714	0.7712	1.5251	RbfSVR	0.7658	0.7654	1.5474
DTR	0.9998	0.9998	0.0393	DTR	0.7793	0.7789	1.5021
RFR	0.9772	0.9772	0.4814	RFR	0.8890	0.8887	1.0655
AdaBoostR	0.5887	0.5883	2.0459	AdaBoostR	0.5814	0.5806	2.0688
GBR	0.7389	0.7387	1.6300	GBR	0.7327	0.7322	1.6532
KNNR	0.8722	0.8721	1.1405	KNNR	0.7856	0.7851	1.4808
MLPR	0.8056	0.8054	1.4233	MLPR	0.8068	0.8064	1.4056

Panel D: ESG Issues

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5686	0.5678	2.0952	Linear	0.5704	0.5685	2.0959
ElasticNet	0.4431	0.4420	2.3806	ElasticNet	0.4429	0.4404	2.3868
Ridge	0.5686	0.5678	2.0953	Ridge	0.5704	0.5685	2.0959
OrdinalRidge	0.5619	0.5611	2.1114	OrdinalRidge	0.5639	0.5620	2.1117
LinearSVR	0.5611	0.5603	2.1134	LinearSVR	0.5632	0.5613	2.1133
RbfSVR	0.7134	0.7129	1.7078	RbfSVR	0.6919	0.6906	1.7749
DTR	0.9994	0.9994	0.0800	DTR	0.7767	0.7758	1.5109
RFR	0.9771	0.9771	0.4825	RFR	0.8873	0.8868	1.0737
AdaBoostR	0.5842	0.5834	2.0571	AdaBoostR	0.5749	0.5730	2.0848
GBR	0.7360	0.7355	1.6390	GBR	0.7298	0.7286	1.6622
KNNR	0.8933	0.8931	1.0419	KNNR	0.8217	0.8209	1.3502
MLPR	0.8104	0.8100	1.3892	MLPR	0.7636	0.7626	1.5547

Notes: This table displays the robustness results using an increased test sample size. The table reports the three most common measures of prediction accuracy for all regression models: R^2, adj. R^2, and root mean squared error (RMSE). In Panel A, we used the baseline specification and thus no CSI predictor. In Panel B, we used Current RRI as a predictor to capture CSI exposure. In Panel C, we used Peak RRI as a predictor to capture CSI exposure. In Panel D, we used the vector of ESG issues as a predictor to capture CSI exposure. The vector of additional accounting- and market-related variables and the vector of sector dummies were included in all models. For more information on algorithm specifications, please see Table C.13.

C.7 Robustness: Alternative Specification

Two of the Altman (1968) variables, OPEF and AT, raised some collinearity concerns with correlations larger than .8 with other predictors. In an alternative specification, and similar to Barboza et al. (2017), we dropped OPEF and AT from our models. Our results are robust when considering this alternative specification. The prediction accuracy of all our models neither increases nor decreases considerably after removing the two variables with high collinearity (cf., Table C.17).

Table C.17: Robustness Check – Alternative Specification

Panel A: Baseline – Without CSI Exposure

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5587	0.5584	2.1220	Linear	0.5624	0.5612	2.1067
ElasticNet	0.4416	0.4413	2.3868	ElasticNet	0.4508	0.4494	2.3600
Ridge	0.5587	0.5584	2.1220	Ridge	0.5624	0.5612	2.1067
OrdinalRidge	0.5513	0.5510	2.1395	OrdinalRidge	0.5556	0.5544	2.1229
LinearSVR	0.5513	0.5510	2.1396	LinearSVR	0.5531	0.5519	2.1289
RbfSVR	0.7487	0.7485	1.6014	RbfSVR	0.7545	0.7539	1.5778
DTR	0.9990	0.9990	0.1031	DTR	0.7950	0.7945	1.4417
RFR	0.9787	0.9786	0.4667	RFR	0.8968	0.8965	1.0231
AdaBoostR	0.5734	0.5731	2.0863	AdaBoostR	0.5604	0.5592	2.1115
GBR	0.7310	0.7308	1.6566	GBR	0.7302	0.7295	1.6540
KNNR	0.9076	0.9075	0.9711	KNNR	0.8493	0.8489	1.2364
MLPR	0.7751	0.7750	1.5147	MLPR	0.7861	0.7855	1.4730

Panel B: Current RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5697	0.5694	2.0953	Linear	0.5739	0.5727	2.0788
ElasticNet	0.4501	0.4497	2.3687	ElasticNet	0.4590	0.4575	2.3423
Ridge	0.5697	0.5694	2.0953	Ridge	0.5739	0.5727	2.0788
OrdinalRidge	0.5629	0.5626	2.1118	OrdinalRidge	0.5685	0.5673	2.0919
LinearSVR	0.5634	0.5631	2.1107	LinearSVR	0.5665	0.5653	2.0968
RbfSVR	0.7551	0.7550	1.5806	RbfSVR	0.7612	0.7605	1.5563
DTR	0.9998	0.9998	0.0457	DTR	0.7990	0.7984	1.4277
RFR	0.9788	0.9788	0.4653	RFR	0.8941	0.8938	1.0361
AdaBoostR	0.5875	0.5873	2.0514	AdaBoostR	0.5788	0.5776	2.0668
GBR	0.7303	0.7301	1.6589	GBR	0.7309	0.7302	1.6519
KNNR	0.8125	0.8123	1.3833	KNNR	0.6991	0.6983	1.7468
MLPR	0.7818	0.7816	1.4922	MLPR	0.7878	0.7872	1.4669

Table C.17 continued from the previous page

Panel C: Peak RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5680	0.5677	2.0994	Linear	0.5712	0.5700	2.0852
ElasticNet	0.4487	0.4483	2.3717	ElasticNet	0.4562	0.4547	2.3484
Ridge	0.5680	0.5677	2.0994	Ridge	0.5712	0.5700	2.0853
OrdinalRidge	0.5606	0.5603	2.1173	OrdinalRidge	0.5654	0.5642	2.0994
LinearSVR	0.5615	0.5612	2.1151	LinearSVR	0.5638	0.5626	2.1032
RbfSVR	0.7562	0.7561	1.5770	RbfSVR	0.7624	0.7617	1.5523
DTR	0.9998	0.9998	0.0489	DTR	0.8042	0.8036	1.4092
RFR	0.9794	0.9794	0.4583	RFR	0.8960	0.8957	1.0269
AdaBoostR	0.5777	0.5774	2.0758	AdaBoostR	0.5701	0.5689	2.0880
GBR	0.7335	0.7333	1.6489	GBR	0.7341	0.7333	1.6422
KNNR	0.8811	0.8811	1.1012	KNNR	0.8066	0.8061	1.4003
MLPR	0.7911	0.7909	1.4822	MLPR	0.7987	0.7981	1.4289

Panel D: ESG Issues

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5667	0.5660	2.1027	Linear	0.5708	0.5681	2.0863
ElasticNet	0.4416	0.4407	2.3868	ElasticNet	0.4508	0.4473	2.3600
Ridge	0.5667	0.5660	2.1027	Ridge	0.5708	0.5681	2.0862
OrdinalRidge	0.5604	0.5597	2.1178	OrdinalRidge	0.5650	0.5623	2.1003
LinearSVR	0.5592	0.5585	2.1207	LinearSVR	0.5613	0.5585	2.1094
RbfSVR	0.7044	0.7040	1.7366	RbfSVR	0.6919	0.6900	1.7676
DTR	0.9990	0.9990	0.0988	DTR	0.7961	0.7948	1.4380
RFR	0.9784	0.9783	0.4698	RFR	0.8938	0.8931	1.0378
AdaBoostR	0.5833	0.5826	2.0620	AdaBoostR	0.5716	0.5688	2.0845
GBR	0.7304	0.7299	1.6586	GBR	0.7306	0.7289	1.6529
KNNR	0.8983	0.8982	1.0185	KNNR	0.8349	0.8339	1.2938
MLPR	0.7767	0.7764	1.5093	MLPR	0.7620	0.7605	1.5534

Notes: This table displays the robustness results using an alternative specification accounting for potential collinearity. The table reports the three most common measures of prediction accuracy for all regression models: R^2, adj. R^2, and root mean squared error (RMSE). In Panel A, we used the baseline specification and thus no CSI predictor. In Panel B, we used Current RRI as a predictor to capture CSI exposure. In Panel C, we used Peak RRI as a predictor to capture CSI exposure. In Panel D, we used the vector of ESG issues as a predictor to capture CSI exposure. The vector of additional accounting- and market-related variables and the vector of sector dummies were included in all models. For more information on algorithm specifications, please see Table C.13.

C.8 Robustness: Non-Stratified Samples

Instead of stratifying our sample as in our main analysis, we used nonstratified samples (i.e., the class distribution between the training and test samples being completely random). We did not use nonstratified samples in our main analysis because this may have led to highly imbalanced class distribution in our target variable. For instance, highly rated firms might be overrepresented in the test sample compared to the training sample, or vice versa. A disproportionate distribution of class labels between the training and test samples could adversely affect the prediction accuracy of our models. However, after all, using nonstratified samples did not impair the prediction accuracy of our models (cf., Table C.18). Our results remain robust in terms of a potentially uneven distribution of class labels among the training and test samples.

Table C.18: Robustness Check – Non-Stratified Samples

Panel A: Baseline – Without CSI Exposure

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5627	0.5624	2.1097	Linear	0.5547	0.5534	2.1356
ElasticNet	0.4443	0.4439	2.3784	ElasticNet	0.4445	0.4429	2.3852
Ridge	0.5627	0.5624	2.1097	Ridge	0.5547	0.5534	2.1356
OrdinalRidge	0.5552	0.5548	2.1279	OrdinalRidge	0.5465	0.5452	2.1553
LinearSVR	0.5555	0.5552	2.1270	LinearSVR	0.5484	0.5471	2.1506
RbfSVR	0.7695	0.7693	1.5318	RbfSVR	0.7552	0.7545	1.5836
DTR	0.9991	0.9991	0.0938	DTR	0.7897	0.7890	1.4678
RFR	0.9791	0.9791	0.4611	RFR	0.8840	0.8837	1.0899
AdaBoostR	0.5975	0.5972	2.0241	AdaBoostR	0.5776	0.5764	2.0799
GBR	0.7399	0.7397	1.6270	GBR	0.7212	0.7204	1.6899
KNNR	0.9121	0.9120	0.9461	KNNR	0.8449	0.8444	1.2605
MLPR	0.7932	0.7931	1.4507	MLPR	0.7818	0.7812	1.4950

Panel B: Current RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5739	0.5736	2.0825	Linear	0.5670	0.5657	2.1060
ElasticNet	0.4525	0.4521	2.3607	ElasticNet	0.4551	0.4534	2.3625
Ridge	0.5739	0.5736	2.0825	Ridge	0.5670	0.5657	2.1060
OrdinalRidge	0.5664	0.5660	2.1010	OrdinalRidge	0.5584	0.5571	2.1268
LinearSVR	0.5674	0.5671	2.0984	LinearSVR	0.5612	0.5599	2.1200
RbfSVR	0.7772	0.7770	1.5060	RbfSVR	0.7613	0.7606	1.5635
DTR	0.9998	0.9998	0.0427	DTR	0.7900	0.7894	1.4665
RFR	0.9794	0.9794	0.4582	RFR	0.8830	0.8827	1.0945
AdaBoostR	0.5899	0.5896	2.0430	AdaBoostR	0.5732	0.5719	2.0909
GBR	0.7414	0.7412	1.6224	GBR	0.7228	0.7220	1.6848
KNNR	0.8140	0.8139	1.3759	KNNR	0.6956	0.6946	1.7658
MLPR	0.8164	0.8163	1.3670	MLPR	0.8020	0.8014	1.4241

Table C.18 continued from the previous page

Panel C: Peak RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5724	0.5721	2.0862	Linear	0.5630	0.5617	2.1157
ElasticNet	0.4514	0.4510	2.3632	ElasticNet	0.4510	0.4494	2.3713
Ridge	0.5724	0.5721	2.0862	Ridge	0.5630	0.5617	2.1157
OrdinalRidge	0.5643	0.5640	2.1060	OrdinalRidge	0.5540	0.5527	2.1372
LinearSVR	0.5659	0.5656	2.1021	LinearSVR	0.5574	0.5561	2.1292
RbfSVR	0.7777	0.7775	1.5042	RbfSVR	0.7612	0.7604	1.5641
DTR	0.9998	0.9998	0.0401	DTR	0.7919	0.7913	1.4599
RFR	0.9797	0.9797	0.4541	RFR	0.8860	0.8856	1.0806
AdaBoostR	0.6067	0.6064	2.0008	AdaBoostR	0.5885	0.5873	2.0529
GBR	0.7410	0.7408	1.6238	GBR	0.7225	0.7217	1.6859
KNNR	0.8839	0.8838	1.0871	KNNR	0.8005	0.7999	1.4293
MLPR	0.8124	0.8123	1.4007	MLPR	0.8010	0.8004	1.4278

Panel D: ESG Issues

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5712	0.5704	2.0893	Linear	0.5619	0.5590	2.1183
ElasticNet	0.4443	0.4434	2.3784	ElasticNet	0.4445	0.4409	2.3852
Ridge	0.5711	0.5704	2.0893	Ridge	0.5619	0.5590	2.1183
OrdinalRidge	0.5642	0.5635	2.1062	OrdinalRidge	0.5553	0.5523	2.1342
LinearSVR	0.5639	0.5632	2.1069	LinearSVR	0.5552	0.5523	2.1344
RbfSVR	0.7182	0.7177	1.6937	RbfSVR	0.6875	0.6854	1.7892
DTR	0.9993	0.9993	0.0868	DTR	0.8002	0.7988	1.4307
RFR	0.9790	0.9790	0.4624	RFR	0.8846	0.8839	1.0870
AdaBoostR	0.5892	0.5885	2.0449	AdaBoostR	0.5700	0.5672	2.0985
GBR	0.7388	0.7384	1.6305	GBR	0.7210	0.7191	1.6905
KNNR	0.9028	0.9026	0.9947	KNNR	0.8316	0.8305	1.3133
MLPR	0.8004	0.8000	1.4256	MLPR	0.7642	0.7627	1.5540

Notes: This table displays the robustness results using non-stratified samples. The table reports the three most common measures of prediction accuracy for all regression models: R^2, adj. R^2, and root mean squared error (RMSE). In Panel A, we used the baseline specification and thus no CSI predictor. In Panel B, we used Current RRI as a predictor to capture CSI exposure. In Panel C, we used Peak RRI as a predictor to capture CSI exposure. In Panel D, we used the vector of ESG issues as a predictor to capture CSI exposure. The vector of additional accounting- and market-related variables and the vector of sector dummies were included in all models. For more information on algorithm specifications, please see Table C.13.

C.9 Additional Analysis: Four-Quarter Lag

Table C.19: Additional Analysis – Four Quarters Lag

Panel A: Baseline – Without CSI Exposure							
	Training Sample				Test Sample		
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5360	0.5357	2.1941	Linear	0.4965	0.4950	2.2612
ElasticNet	0.4334	0.4330	2.4246	ElasticNet	0.4289	0.4271	2.4083
Ridge	0.5360	0.5357	2.1941	Ridge	0.4971	0.4955	2.2599
OrdinalRidge	0.5312	0.5308	2.2056	OrdinalRidge	0.5255	0.5240	2.1952
LinearSVR	0.5209	0.5205	2.2297	LinearSVR	0.5027	0.5011	2.2473
RbfSVR	0.7488	0.7486	1.6144	RbfSVR	0.7343	0.7335	1.6425
DTR	0.9989	0.9989	0.1059	DTR	0.7758	0.7751	1.5089
RFR	0.9857	0.9857	0.3848	RFR	0.8961	0.8958	1.0271
AdaBoostR	0.5148	0.5144	2.2438	AdaBoostR	0.5094	0.5078	2.2322
GBR	0.7250	0.7248	1.6892	GBR	0.7026	0.7016	1.7380
KNNR	0.9106	0.9105	0.9632	KNNR	0.8460	0.8456	1.2504
MLPR	0.7904	0.7902	1.4747	MLPR	0.6571	0.6560	1.8661

Panel B: Current RRI							
	Training Sample				Test Sample		
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5458	0.5455	2.1708	Linear	0.5087	0.5072	2.2336
ElasticNet	0.4400	0.4395	2.4105	ElasticNet	0.4376	0.4357	2.3899
Ridge	0.5458	0.5455	2.1708	Ridge	0.5093	0.5077	2.2324
OrdinalRidge	0.5406	0.5403	2.1832	OrdinalRidge	0.5378	0.5363	2.1665
LinearSVR	0.5326	0.5322	2.2023	LinearSVR	0.5183	0.5167	2.2118
RbfSVR	0.7555	0.7553	1.5927	RbfSVR	0.7401	0.7393	1.6246
DTR	0.9998	0.9998	0.0461	DTR	0.7826	0.7819	1.4859
RFR	0.9863	0.9863	0.3770	RFR	0.8964	0.8960	1.0259
AdaBoostR	0.5216	0.5213	2.2279	AdaBoostR	0.5163	0.5147	2.2164
GBR	0.7263	0.7261	1.6852	GBR	0.7044	0.7035	1.7325
KNNR	0.8132	0.8131	1.3920	KNNR	0.6969	0.6959	1.7545
MLPR	0.7807	0.7805	1.5085	MLPR	0.7478	0.7470	1.6003

Table C.19 continued from the previous page

Panel C: Peak RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5445	0.5441	2.1740	Linear	0.5042	0.5026	2.2439
ElasticNet	0.4391	0.4387	2.4124	ElasticNet	0.4354	0.4336	2.3945
Ridge	0.5445	0.5441	2.1740	Ridge	0.5048	0.5032	2.2425
OrdinalRidge	0.5402	0.5398	2.1843	OrdinalRidge	0.5344	0.5329	2.1745
LinearSVR	0.5306	0.5302	2.2069	LinearSVR	0.5101	0.5086	2.2304
RbfSVR	0.7561	0.7559	1.5908	RbfSVR	0.7398	0.7390	1.6255
DTR	0.9997	0.9997	0.0563	DTR	0.7900	0.7893	1.4604
RFR	0.9865	0.9865	0.3738	RFR	0.8971	0.8968	1.0220
AdaBoostR	0.5273	0.5269	2.2146	AdaBoostR	0.5200	0.5185	2.2078
GBR	0.7291	0.7289	1.6765	GBR	0.7072	0.7063	1.7244
KNNR	0.8816	0.8815	1.1082	KNNR	0.7922	0.7915	1.4527
MLPR	0.8062	0.8060	1.4367	MLPR	0.7547	0.7539	1.5784

Panel D: ESG Issues

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5433	0.5425	2.1768	Linear	0.5013	0.4979	2.2504
ElasticNet	0.4334	0.4324	2.4246	ElasticNet	0.4289	0.4249	2.4083
Ridge	0.5433	0.5425	2.1769	Ridge	0.5019	0.4984	2.2491
OrdinalRidge	0.5386	0.5378	2.1879	OrdinalRidge	0.5292	0.5259	2.1867
LinearSVR	0.5279	0.5270	2.2133	LinearSVR	0.5055	0.5021	2.2409
RbfSVR	0.6951	0.6945	1.7788	RbfSVR	0.6683	0.6660	1.8353
DTR	0.9991	0.9991	0.0968	DTR	0.7836	0.7821	1.4824
RFR	0.9857	0.9857	0.3853	RFR	0.8945	0.8938	1.0350
AdaBoostR	0.5218	0.5210	2.2274	AdaBoostR	0.5152	0.5118	2.2189
GBR	0.7248	0.7243	1.6898	GBR	0.7024	0.7004	1.7384
KNNR	0.9004	0.9003	1.0164	KNNR	0.8295	0.8283	1.3160
MLPR	0.7879	0.7875	1.4834	MLPR	0.7159	0.7139	1.6986

Notes: This table displays the robustness results using a four- instead of two-quarter lag for the predictor variables. The table reports the three most common measures of prediction accuracy for all regression models: R^2, adj. R^2, and root mean squared error (RMSE). In Panel A, we used the baseline specification and thus no CSI predictor. In Panel B, we used Current RRI as a predictor to capture CSI exposure. In Panel C, we used Peak RRI as a predictor to capture CSI exposure. In Panel D, we used the vector of ESG issues as a predictor to capture CSI exposure. The vector of additional accounting- and market-related variables and the vector of sector dummies were included in all models. For more information on algorithm specifications, please see Table C.13.

C.10 Additional Analysis: Post-2015 Data

Table C.20: Additional Analysis – Post-2015 Subsample

Panel A: Baseline – Without CSI Exposure

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5730	0.5724	2.0770	Linear	0.5759	0.5736	2.0464
ElasticNet	0.4580	0.4573	2.3399	ElasticNet	0.4448	0.4417	2.3415
Ridge	0.5730	0.5724	2.0770	Ridge	0.5757	0.5734	2.0469
OrdinalRidge	0.5654	0.5648	2.0954	OrdinalRidge	0.5693	0.5670	2.0622
LinearSVR	0.5641	0.5635	2.0985	LinearSVR	0.5707	0.5683	2.0590
RbfSVR	0.7753	0.7749	1.5068	RbfSVR	0.7639	0.7626	1.5270
DTR	0.9989	0.9989	0.1043	DTR	0.7980	0.7969	1.4122
RFR	0.9870	0.9870	0.3620	RFR	0.9165	0.9160	0.9081
AdaBoostR	0.6150	0.6145	1.9721	AdaBoostR	0.6094	0.6072	1.9640
GBR	0.7587	0.7583	1.5614	GBR	0.7466	0.7452	1.5818
KNNR	0.9223	0.9222	0.8858	KNNR	0.8713	0.8706	1.1272
MLPR	0.8442	0.8440	1.2545	MLPR	0.8191	0.8181	1.3365

Panel B: Current RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5843	0.5837	2.0493	Linear	0.5871	0.5847	2.0192
ElasticNet	0.4662	0.4655	2.3221	ElasticNet	0.4533	0.4501	2.3235
Ridge	0.5843	0.5837	2.0494	Ridge	0.5869	0.5845	2.0197
OrdinalRidge	0.5759	0.5753	2.0698	OrdinalRidge	0.5824	0.5800	2.0306
LinearSVR	0.5762	0.5756	2.0692	LinearSVR	0.5831	0.5807	2.0290
RbfSVR	0.7836	0.7833	1.4785	RbfSVR	0.7748	0.7735	1.4912
DTR	0.9999	0.9999	0.0314	DTR	0.7921	0.7909	1.4328
RFR	0.9874	0.9874	0.3565	RFR	0.9162	0.9158	0.9094
AdaBoostR	0.6175	0.6170	1.9656	AdaBoostR	0.6090	0.6068	1.9648
GBR	0.7579	0.7575	1.5639	GBR	0.7453	0.7438	1.5860
KNNR	0.8118	0.8116	1.3787	KNNR	0.6887	0.6870	1.7531
MLPR	0.8408	0.8405	1.2684	MLPR	0.8182	0.8172	1.3398

Table C.20 continued from the previous page

Panel C: Peak RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5821	0.5815	2.0546	Linear	0.5854	0.5830	2.0234
ElasticNet	0.4642	0.4635	2.3265	ElasticNet	0.4530	0.4499	2.3240
Ridge	0.5821	0.5815	2.0546	Ridge	0.5852	0.5828	2.0238
OrdinalRidge	0.5747	0.5741	2.0727	OrdinalRidge	0.5779	0.5755	2.0416
LinearSVR	0.5733	0.5727	2.0763	LinearSVR	0.5808	0.5784	2.0345
RbfSVR	0.7840	0.7836	1.4773	RbfSVR	0.7752	0.7739	1.4900
DTR	1.0000	1.0000	0.0000	DTR	0.8024	0.8013	1.3967
RFR	0.9880	0.9880	0.3478	RFR	0.9188	0.9183	0.8956
AdaBoostR	0.6280	0.6275	1.9385	AdaBoostR	0.6213	0.6192	1.9336
GBR	0.7596	0.7593	1.5583	GBR	0.7476	0.7462	1.5785
KNNR	0.8824	0.8822	1.0900	KNNR	0.7871	0.7859	1.4498
MLPR	0.8450	0.8448	1.2820	MLPR	0.8261	0.8251	1.3103

Panel D: ESG Issues

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.5827	0.5815	2.0531	Linear	0.5838	0.5786	2.0273
ElasticNet	0.4580	0.4564	2.3399	ElasticNet	0.4448	0.4379	2.3415
Ridge	0.5827	0.5814	2.0531	Ridge	0.5836	0.5784	2.0277
OrdinalRidge	0.5743	0.5730	2.0737	OrdinalRidge	0.5745	0.5692	2.0498
LinearSVR	0.5733	0.5720	2.0762	LinearSVR	0.5789	0.5737	2.0391
RbfSVR	0.7179	0.7171	1.6881	RbfSVR	0.6964	0.6927	1.7313
DTR	0.9990	0.9990	0.1008	DTR	0.8033	0.8008	1.3938
RFR	0.9869	0.9868	0.3640	RFR	0.9155	0.9144	0.9136
AdaBoostR	0.6271	0.6259	1.9410	AdaBoostR	0.6241	0.6194	1.9265
GBR	0.7575	0.7567	1.5653	GBR	0.7444	0.7412	1.5886
KNNR	0.9113	0.9110	0.9465	KNNR	0.8543	0.8525	1.1996
MLPR	0.8491	0.8486	1.2347	MLPR	0.7288	0.7254	1.6364

Notes: This table displays the robustness results using only post-2015 data. The table reports the three most common measures of prediction accuracy for all regression models: R^2, adj. R^2, and root mean squared error (RMSE). In Panel A, we used the baseline specification and thus no CSI predictor. In Panel B, we used Current RRI as a predictor to capture CSI exposure. In Panel C, we used Peak RRI as a predictor to capture CSI exposure. In Panel D, we used the vector of ESG issues as a predictor to capture CSI exposure. The vector of additional accounting- and market-related variables and the vector of sector dummies were included in all models. For more information on algorithm specifications, please see Table C.13.

C.11 Additional Analysis: Moody's and Fitch Credit Rating Data

Similar to S&P ratings, Moody's and Fitch ratings were converted to numerical values according to the respective scale. Using Moody's and Fitch rating data considerably reduced our sample to 8,202 and 4,261 firm-quarter observations, respectively. Again, the two samples were split into a training (80%) and a test (20%) sample. Table C.21 shows the results of rerunning the analysis with credit risk data from Moody's. Table C.22 shows the results of rerunning the analysis with credit risk data from Fitch.

The findings suggest that credit risk, as measured by Moody's and Fitch credit ratings, is much less explained by more traditional algorithms. Adjusted R^2s decrease considerably to only slightly above 20% for most models (except when using ElasticNet, which is even well-below 20%). While most of the more advanced ML algorithms maintain their comparably high out-of-sample prediction accuracy, only Random Forest regression, K-nearest Neighbor regression, and Multi-layer Perceptron regression achieve outstanding out-of-sample prediction accuracy close to or above 80% for adj. R^2. This finding is somewhat surprising since algorithms like Random Forest regression usually work best with sample sizes well-beyond 10,000. In our case, and despite the considerably decreased sample size, the mentioned algorithms maintain their prediction accuracy scores.

Table C.21: Additional Analysis – Moody's Credit Rating Data

Panel A: Baseline – Without CSI Exposure

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.2739	0.2712	1.6982	Linear	0.2439	0.2322	1.7128
ElasticNet	0.1279	0.1246	1.8611	ElasticNet	0.1202	0.1066	1.8477
Ridge	0.2726	0.2698	1.6998	Ridge	0.2421	0.2304	1.7149
OrdinalRidge	0.2568	0.2540	1.7181	OrdinalRidge	0.2265	0.2146	1.7324
LinearSVR	0.2604	0.2576	1.7139	LinearSVR	0.2245	0.2125	1.7347
RbfSVR	0.6794	0.6781	1.1285	RbfSVR	0.6533	0.6479	1.1598
DTR	0.9992	0.9992	0.0559	DTR	0.7450	0.7410	0.9948
RFR	0.9745	0.9744	0.3181	RFR	0.8581	0.8560	0.7419
AdaBoostR	0.3794	0.3770	1.5700	AdaBoostR	0.3624	0.3525	1.5730
GBR	0.6732	0.6720	1.1392	GBR	0.6465	0.6410	1.1712
KNNR	0.9018	0.9014	0.6247	KNNR	0.8447	0.8423	0.7763
MLPR	0.8561	0.8555	0.7561	MLPR	0.7851	0.7817	0.9132

Panel B: Current RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.3104	0.3077	1.6550	Linear	0.2874	0.2759	1.6629
ElasticNet	0.1606	0.1573	1.8259	ElasticNet	0.1627	0.1492	1.8025
Ridge	0.3093	0.3066	1.6563	Ridge	0.2862	0.2747	1.6642
OrdinalRidge	0.2900	0.2872	1.6793	OrdinalRidge	0.2677	0.2559	1.6857
LinearSVR	0.2978	0.2950	1.6700	LinearSVR	0.2704	0.2586	1.6826
RbfSVR	0.6918	0.6906	1.1064	RbfSVR	0.6627	0.6573	1.1440
DTR	0.9999	0.9999	0.0195	DTR	0.6931	0.6882	1.0912
RFR	0.9747	0.9746	0.3170	RFR	0.8611	0.8589	0.7341
AdaBoostR	0.3658	0.3632	1.5872	AdaBoostR	0.3464	0.3359	1.5925
GBR	0.6855	0.6842	1.1177	GBR	0.6465	0.6409	1.1711
KNNR	0.7147	0.7136	1.0644	KNNR	0.5610	0.5539	1.3051
MLPR	0.8895	0.8891	0.6624	MLPR	0.8206	0.8177	0.8343

Table C.21 continued from the previous page

Panel C: Peak RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.6868	0.6865	1.7952	Linear	0.6942	0.6928	1.7607
ElasticNet	0.5398	0.5393	2.1760	ElasticNet	0.5404	0.5383	2.1585
Ridge	0.6868	0.6864	1.7953	Ridge	0.6940	0.6926	1.7613
OrdinalRidge	0.6795	0.6792	1.8159	OrdinalRidge	0.6865	0.6851	1.7826
LinearSVR	0.6803	0.6799	1.8138	LinearSVR	0.6867	0.6853	1.7821
RbfSVR	0.8519	0.8518	1.2344	RbfSVR	0.8432	0.8425	1.2606
DTR	0.9998	0.9998	0.0409	DTR	0.8333	0.8325	1.3001
RFR	0.9856	0.9855	0.3855	RFR	0.9256	0.9253	0.8682
AdaBoostR	0.6640	0.6636	1.8595	AdaBoostR	0.6529	0.6513	1.8758
GBR	0.8214	0.8212	1.3555	GBR	0.8127	0.8119	1.3778
KNNR	0.9011	0.9010	1.0089	KNNR	0.8277	0.8269	1.3217
MLPR	0.9043	0.9042	1.0179	MLPR	0.8849	0.8844	1.0800

Panel D: ESG Issues

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.6896	0.6889	1.7871	Linear	0.6955	0.6925	1.7570
ElasticNet	0.5451	0.5440	2.1635	ElasticNet	0.5469	0.5424	2.1433
Ridge	0.6896	0.6888	1.7873	Ridge	0.6952	0.6923	1.7577
OrdinalRidge	0.6822	0.6814	1.8085	OrdinalRidge	0.6890	0.6860	1.7755
LinearSVR	0.6827	0.6819	1.8070	LinearSVR	0.6880	0.6849	1.7785
RbfSVR	0.7910	0.7904	1.4666	RbfSVR	0.7697	0.7675	1.5278
DTR	0.9989	0.9989	0.1058	DTR	0.8342	0.8326	1.2963
RFR	0.9845	0.9845	0.3989	RFR	0.9241	0.9234	0.8770
AdaBoostR	0.6759	0.6751	1.8263	AdaBoostR	0.6705	0.6673	1.8277
GBR	0.8216	0.8211	1.3550	GBR	0.8130	0.8111	1.3770
KNNR	0.9252	0.9250	0.8773	KNNR	0.8792	0.8780	1.1068
MLPR	0.8986	0.8983	1.0216	MLPR	0.8259	0.8242	1.3284

Notes: This table displays the robustness results using credit rating data from Moody's. The table reports the three most common measures of prediction accuracy for all regression models: R^2, adj. R^2, and root mean squared error (RMSE). In Panel A, we used the baseline specification and thus no CSI predictor. In Panel B, we used Current RRI as a predictor to capture CSI exposure. In Panel C, we used Peak RRI as a predictor to capture CSI exposure. In Panel D, we used the vector of ESG issues as a predictor to capture CSI exposure. The vector of additional accounting- and market-related variables and the vector of sector dummies were included in all models. For more information on algorithm specifications, please see Table C.13.

Table C.22: Additional Analysis – Fitch Credit Rating Data

Panel A: Baseline – Without CSI Exposure

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.4700	0.4661	2.0764	Linear	0.4147	0.3970	2.1609
ElasticNet	0.3571	0.3524	2.2868	ElasticNet	0.3543	0.3347	2.2697
Ridge	0.4680	0.4641	2.0802	Ridge	0.4118	0.3941	2.1662
OrdinalRidge	0.4577	0.4537	2.1004	OrdinalRidge	0.4085	0.3907	2.1722
LinearSVR	0.4540	0.4500	2.1074	LinearSVR	0.4032	0.3851	2.1821
RbfSVR	0.7434	0.7415	1.4447	RbfSVR	0.6603	0.6501	1.6462
DTR	1.0000	1.0000	0.0000	DTR	0.7778	0.7711	1.3314
RFR	0.9691	0.9689	0.5015	RFR	0.8340	0.8289	1.1510
AdaBoostR	0.5085	0.5049	1.9995	AdaBoostR	0.4555	0.4391	2.0842
GBR	0.7741	0.7724	1.3557	GBR	0.6822	0.6726	1.5923
KNNR	0.9201	0.9195	0.8062	KNNR	0.8819	0.8783	0.9708
MLPR	0.9219	0.9213	0.7973	MLPR	0.8519	0.8474	1.0871

Panel B: Current RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.4755	0.4714	2.0657	Linear	0.4273	0.4093	2.1376
ElasticNet	0.3649	0.3600	2.2730	ElasticNet	0.3703	0.3505	2.2413
Ridge	0.4738	0.4698	2.0689	Ridge	0.4253	0.4072	2.1412
OrdinalRidge	0.4638	0.4596	2.0886	OrdinalRidge	0.4224	0.4042	2.1467
LinearSVR	0.4604	0.4562	2.0952	LinearSVR	0.4166	0.3982	2.1575
RbfSVR	0.7533	0.7514	1.4167	RbfSVR	0.6735	0.6632	1.6139
DTR	1.0000	1.0000	0.0000	DTR	0.7696	0.7623	1.3558
RFR	0.9698	0.9695	0.4960	RFR	0.8252	0.8197	1.1810
AdaBoostR	0.5086	0.5048	1.9994	AdaBoostR	0.4574	0.4403	2.0806
GBR	0.7713	0.7695	1.3641	GBR	0.6784	0.6683	1.6018
KNNR	0.7731	0.7713	1.3586	KNNR	0.6758	0.6656	1.6083
MLPR	0.9134	0.9127	0.8394	MLPR	0.8516	0.8469	1.0882

Table C.22 continued from the previous page

Panel C: Peak RRI

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.4798	0.4758	2.0571	Linear	0.4310	0.4131	2.1306
ElasticNet	0.3698	0.3649	2.2642	ElasticNet	0.3734	0.3536	2.2359
Ridge	0.4783	0.4743	2.0600	Ridge	0.4292	0.4112	2.1340
OrdinalRidge	0.4678	0.4637	2.0807	OrdinalRidge	0.4119	0.3934	2.1660
LinearSVR	0.4654	0.4612	2.0855	LinearSVR	0.4223	0.4041	2.1469
RbfSVR	0.7581	0.7563	1.4027	RbfSVR	0.6738	0.6636	1.6131
DTR	1.0000	1.0000	0.0000	DTR	0.7495	0.7416	1.4138
RFR	0.9702	0.9700	0.4920	RFR	0.8379	0.8328	1.1371
AdaBoostR	0.5215	0.5178	1.9730	AdaBoostR	0.4562	0.4391	2.0828
GBR	0.7787	0.7770	1.3416	GBR	0.6847	0.6747	1.5861
KNNR	0.8616	0.8606	1.0609	KNNR	0.7916	0.7850	1.2894
MLPR	0.9424	0.9420	0.8180	MLPR	0.8641	0.8598	1.0413

Panel D: ESG Issues

Training Sample				Test Sample			
Algorithm	R^2	adj. R^2	RMSE	Algorithm	R^2	adj. R^2	RMSE
Linear	0.4792	0.4705	2.0582	Linear	0.4118	0.3704	2.1663
ElasticNet	0.3571	0.3464	2.2868	ElasticNet	0.3543	0.3088	2.2697
Ridge	0.4771	0.4684	2.0624	Ridge	0.4121	0.3708	2.1656
OrdinalRidge	0.4692	0.4604	2.0779	OrdinalRidge	0.4121	0.3707	2.1658
LinearSVR	0.4638	0.4548	2.0886	LinearSVR	0.4000	0.3578	2.1879
RbfSVR	0.6306	0.6244	1.7335	RbfSVR	0.5322	0.4993	1.9318
DTR	1.0000	1.0000	0.0000	DTR	0.7461	0.7282	1.4233
RFR	0.9686	0.9681	0.5051	RFR	0.8306	0.8187	1.1626
AdaBoostR	0.5185	0.5105	1.9791	AdaBoostR	0.4559	0.4176	2.0835
GBR	0.7740	0.7702	1.3559	GBR	0.6764	0.6536	1.6069
KNNR	0.9110	0.9095	0.8509	KNNR	0.8635	0.8539	1.0437
MLPR	0.9136	0.9122	0.8382	MLPR	0.6988	0.6776	1.5501

Notes: This table displays the robustness results using credit rating data from Fitch. The table reports the three most common measures of prediction accuracy for all regression models: R^2, adj. R^2, and root mean squared error (RMSE). In Panel A, we used the baseline specification and thus no CSI predictor. In Panel B, we used Current RRI as a predictor to capture CSI exposure. In Panel C, we used Peak RRI as a predictor to capture CSI exposure. In Panel D, we used the vector of ESG issues as a predictor to capture CSI exposure. The vector of additional accounting- and market-related variables and the vector of sector dummies were included in all models. For more information on algorithm specifications, please see Table C.13.

DANIEL V. FAUSER

EDUCATION

University of St. Gallen (HSG) *September 2017 - August 2020*
Ph.D. in Management (PMA)
School of Management (SoM)

Maastricht University (UM) *May 2019 - September 2019*
Research Visit
Finance Department at the School of Business and Economics (SBE)

University of St. Gallen (HSG) *September 2015 - August 2017*
Master (M.A.) in Accounting and Finance (MAccFin)

University of St. Gallen (HSG) *September 2012 - August 2015*
Bachelor (B.A.) in Business Administration

WORK EXPERIENCE

Swisscanto Invest by Zürcher Kantonalbank Since September 2020
Equity Analyst - Global ESG Integrated

University of St.Gallen (HSG) September 2015 - August 2020
Research and Teaching Assistant

University of St.Gallen (HSG) July 2015 - November 2015
Student Assistant

Kaiser Partner July 2013 - February 2015
Trainee

TECHNICAL STRENGTHS

Programming & Software Python, HTML, LaTeX, Excel, Stata

POSITION OF RESPONSIBILITY - EXTRACURRICULAR

Doctoral Network at the University of St. Gallen (DocNet)
February 2019 - January 2020
Board Member - Sponsoring *University of St. Gallen (HSG)*

Doctoral Network at the University of St. Gallen (DocNet)
February 2018 - January 2019
Board Member - Events *University of St. Gallen (HSG)*

Student Parliament, Program and Finance Commission
September 2017 - August 2018
PMA Representative *University of St. Gallen (HSG)*

Mentoring Program
September 2017 - August 2019
Mentee *University of St. Gallen (HSG)*

sneep St. Gallen
May 2015 - February 2019
Board Member - Finance *University of St. Gallen (HSG)*

HSG Unisport
September 2013 - August 2020
Social Media Advisor / Badminton Coach *University of St. Gallen (HSG)*