INSTITUTO
UNIVERSITÁRIO
DE LISBOA

ESG Controversies and Global Equity Returns: An Event Study Across Different Time Horizons

Nic Kugies

Master Degree in Management

Supervisor:
PhD, Luciana Salles Barbosa, Assistant Professor
Iscte – Instituto Universitário de Lisboa

Co-Supervisor:
PhD, Andrea Sofia Meireles Rodrigues, Assistant Professor
Iscte – Instituto Universitário de Lisboa

May, 2026

BUSINESS
SCHOOL

Department of Marketing, Operations and General Management

ESG Controversies and Global Equity Returns: An Event Study Across Different Time Horizons

Nic Kugies

Master Degree in Management

Supervisor:
PhD, Luciana Salles Barbosa, Assistant Professor
Iscte – Instituto Universitário de Lisboa

Co-Supervisor:
PhD, Andrea Sofia Meireles Rodrigues, Assistant Professor
Iscte – Instituto Universitário de Lisboa

May, 2026

Resumo

As controvérsias ESG — alegações ou incidentes publicamente divulgados de má conduta ambiental, social ou de governança — podem desencadear choques reputacionais e regulatórios e, por isso, ser incorporadas nos preços pelos mercados de capitais. No entanto, a evidência existente é mista quanto à magnitude e à persistência das reações dos preços das ações, bem como quanto à possibilidade de os mercados diferenciarem entre pilares ESG e características das empresas. Utilizando 114 eventos de controvérsias ESG da Refinitiv que afetaram empresas cotadas em bolsa a nível mundial entre 2015 e 2024, realizo um estudo de eventos de curto prazo. Os retornos anormais são estimados com um modelo de mercado ao longo de uma janela de estimação de [-250, -20] dias de negociação e agregados em retornos anormais acumulados ao longo de três horizontes de evento: um horizonte curto [-1, +1], um horizonte intermédio [-1, +5] e um horizonte curto alargado [-1, +20]. Complemento os testes do estudo de eventos com análises transversais para examinar a heterogeneidade por pilar e por proxies de resiliência ao nível da empresa. As divulgações de controvérsias desencadeiam perdas de valor economicamente elevadas e estatisticamente robustas: o CAAR médio é de aproximadamente -12% em [-1, +1], aprofunda-se para cerca de -19% em [-1, +5] e mantém-se fortemente negativo, em torno de -19%, em [-1, +20]. As controvérsias de governança geram as penalizações mais severas, em linha com um canal de credibilidade e de controlo interno. As perdas são atenuadas no caso de empresas com uma reputação ESG pré-evento mais forte e maior capacidade, em particular em termos de dimensão e rentabilidade. Esta tese contribui ao combinar comparações entre pilares e moderadores ao nível da empresa numa amostra multicountry validada manualmente, clarificando quando e por que razão as controvérsias ESG são mais destrutivas em termos de valor.

Abstract

ESG controversies—publicly disclosed allegations or incidents of environmental, social, or governance misconduct—can trigger reputational and regulatory shocks and may therefore be priced by capital markets. Yet existing evidence is mixed on the magnitude and persistence of stock-price reactions and on whether markets differentiate across ESG pillars and firm characteristics. Using 114 Refinitiv ESG Controversies events affecting listed firms worldwide (2015–2024), I conduct a short-horizon event study. Abnormal returns are estimated with a market model over an estimation window of [-250, -20] trading days and aggregated into cumulative abnormal returns across three event horizons: a short horizon [-1, +1], a medium horizon [-1, +5], and a longer short-run horizon [-1, +20]. I complement the event-study tests with cross-sectional analyses to examine heterogeneity by pillar and firm-level resilience proxies. Controversy disclosures trigger economically large and statistically robust value losses: average CAAR is approximately -12% in [-1, +1], deepens to about -19% in [-1, +5], and remains strongly negative at roughly -19% in [-1, +20]. Governance controversies drive the most severe penalties, consistent with a credibility and internal-control channel. Losses are attenuated for firms with stronger pre-event ESG reputation and greater capacity, particularly size and profitability. This thesis contributes by combining pillar comparisons and firm-level moderators within a manually validated, multi-country event sample, clarifying when and why ESG controversies are most value-destructive.

Contents

Resumo i

Abstract iii

List of Tables ix

Glossary of Acronyms xi

Chapter 1. Introduction 1

Chapter 2. Literature Review and Conceptual Framework 5
2.1. Introduction and Concept Definition 5
2.2. Economic Mechanisms: The “Why” 6
2.2.1. Information and Cash-Flow Revision 6
2.2.2. Reputation, Trust, and the Discount-Rate Channel 6
2.2.3. Agency and Governance Signal 7
2.3. Empirical Evidence: The “What” 7
2.3.1. Direction and Magnitude of Reactions 8
2.3.2. Timing: Immediate Versus Lasting Effects 8
2.3.3. Heterogeneity by Pillar, Severity, and Firm Characteristics 9
2.3.4. ESG Ratings as Buffer Versus Backlash: The Greenwashing Link 9
2.4. Research Gap 10
2.5. Hypotheses 10
H1: Baseline Effect 10
H2: Pillar Heterogeneity 11
H3: Moderators of Controversy-Related Value Loss 11

Chapter 3. Methodology 13
3.1. Research Design and Approach 13
3.2. Data and Sample 13
3.2.1. Data Source 13
3.2.2. Sample Selection 14
3.2.3. Stock Returns and Market Data 14
3.3. Operationalization and Variables 15
3.3.1. Normal and Abnormal Returns 15
3.3.2. Cumulative Abnormal Return (CAR) 16
3.3.3. ESG Pillar Classification 16

3.3.4. Firm Characteristics and Dummy Variables 17
3.4. Event-Study Design 19
3.4.1. Estimation Window 19
3.4.2. Event Windows 19
3.4.3. Abnormal Return Calculation 20
3.4.4. Expectations and Sign Convention 20
3.4.5. Multiple Events and Independence 20
3.5. Statistical Tests and Inference 20
3.5.1. Tests for Abnormal Return Significance 21
3.5.2. Non-Parametric Tests 21
3.5.3. Comparing ESG Pillar Reactions 21
3.5.4. Heterogeneity by Firm Characteristics 22
3.6. Robustness Checks 22
3.6.1. Sensitivity to Outliers 22
3.6.2. Temporal Stability 22
3.6.3. Robust Inference in Cross-Sectional Regressions 22

Chapter 4. Empirical Results 23
4.1. Sample Composition & Descriptive Statistics 23
4.1.1. Firm Characteristics 23
4.2. Main Event Study Results 24
4.3. Heterogeneity by ESG Pillar 25
4.4. Moderation Analysis 26
4.4.1. ESG Reputation as a Moderator (H3a) 26
4.4.2. Firm Size as a Moderator (H3b) 26
4.4.3. Financial Health as a Moderator (H3c) 27
4.5. Robustness Checks 29
4.5.1. Outliers and Winsorization 29
4.5.2. Temporal Stability: Pre-2020 vs. 2020–2024 29
4.6. Summary of Findings and Hypothesis Support 30

Chapter 5. Discussion 31
5.1. Main Findings 31
5.2. Comparison with Literature 32
5.3. Theoretical Implications 34
5.4. Practical Implications 36
5.5. Limitations 37
5.6. Future Research 38

Chapter 6. Conclusion 41
6.1. Comprehensive Synthesis of Findings 41
6.2. Theoretical Contributions 42

6.3. Practical and Managerial Implications 43
6.4. Limitations and Future Research 44

References 45

Appendix A. Appendix 49

List of Tables

3.1	List of equations used in the thesis	15
4.1	Descriptive statistics of ESG controversy events	23
4.2	Descriptive statistics of firm characteristics (t–1)	24
4.3	Cumulative average abnormal returns (CAAR) by event window	24
4.4	Cross-sectional regression of CAR (medium window)	25
4.5	CAAR by ESG pillar and event window	26
4.6	ESG reputation and market reaction (high vs. low ESG)	27
4.7	Firm size as a moderator (median split)	27
4.8	Leverage as a moderator (median split)	28
4.9	Profitability as a moderator (median split)	28
4.10	Robustness checks (CAR-based)	29
5.1	Summary of Key Empirical Results (N=114 Events)	31
A.1	Benchmark index mapping by primary listing market	49
A.2	Adjusted inference for CAAR (calendar-date clustered)	50
A.3	Cross-sectional regressions with controls (ESG reputation and firm fundamentals)	50
A.4	Variance inflation factors (VIF) for ESG-control regressions	50

Glossary of Acronyms

AAR: Average abnormal return
ANOVA: Analysis of variance
AR: Abnormal return
CAAR: Cumulative average abnormal return
CAPM: Capital Asset Pricing Model
CAR: Cumulative abnormal return
CEO: Chief executive officer
CSR: Corporate social responsibility
DAX: Deutscher Aktienindex
E: Environmental pillar
ESG: Environmental, Social, and Governance
FTSE 100: Financial Times Stock Exchange 100 Index
G: Governance pillar
LSEG: London Stock Exchange Group
OLS: Ordinary least squares
ROA: Return on assets
S: Social pillar
S&P 500: Standard & Poor's 500 Index
UK: United Kingdom
USD: United States dollar

CHAPTER 1

Introduction

Environmental, social, and governance (ESG) factors have evolved from niche screening criteria into a mainstream element of capital markets (Friede et al., 2015; Berg et al., 2022). Both investors and regulators increasingly view ESG performance as an indicator of a firm's long-term resilience (Khan et al., 2016; Giese et al., 2019). Resilience is defined as a firm's ability to withstand and manage non-financial shocks and risks, as well as the effectiveness of its governance and risk management practices. This shift is reflected in the growth of sustainable investing (Lins et al., 2017). According to the Global Sustainable Investment Alliance, sustainable investment assets amounted to USD 30.3 trillion globally in 2022, indicating that ESG-related information has become part of many investors' standard decision-making processes (Global Sustainable Investment Alliance, 2023).

At the same time, ESG-related controversies have become a recurring feature of corporate news. Environmental accidents, labor and human rights violations, product safety issues, and governance failures, such as fraud or corruption, can quickly undermine stakeholder trust. These incidents can create direct costs, such as legal penalties, clean-up efforts, and compliance investments, as well as indirect costs, including revenue losses, increased regulatory scrutiny, and higher perceived risk. For investors, therefore, controversies represent negative information shocks that can affect both expected cash flows and discount rates.

From a theoretical perspective, ESG controversies can be interpreted as instances of corporate social irresponsibility that threaten legitimacy and reputational capital. According to stakeholder theory, public scrutiny can result in economic consequences when stakeholders reconsider their support for the firm. Media coverage plays a central role in this process by increasing the public's awareness of negative information and accelerating stakeholder responses (Rock, 2022). When controversies become widely visible, firms may face increased pressure from customers, employees, communities, and regulators, amplifying the financial consequences of the initial incident.

Empirically, prior event-study evidence generally points to negative market reactions to adverse ESG news. Studies such as Krüger (2015) and Capelle-Blancard and Petit (2019) document an asymmetric response pattern in which negative ESG news is followed by significant negative abnormal returns, whereas positive ESG news tends to generate smaller or statistically weaker valuation effects. This asymmetry is consistent with the idea that favorable ESG signals may be partially anticipated while controversies reveal downside risks that were not fully priced in.

However, the severity of the market penalty is unlikely to be consistent across firms and types of controversy. Depending on a firm's reputation and prior ESG credibility, investors may

interpret the same incident differently. Aouadi and Marsat (2018) show that the relationship between ESG controversies, corporate social performance, and firm value depends on attention and visibility. This implies that reputation and prominence shape how markets process negative ESG information. This perspective also aligns with the view that ESG controversies are multidimensional since environmental, social, and governance failures differ in their perceived severity, remediability, and implications for internal controls.

Behavioral finance provides an additional mechanism for sharp, short-term reactions. When ESG issues are important, investors may temporarily pay more attention to controversial news than to other information. Consistent with this attention-based channel, Cui and Docherty (2020) argue that controversy announcements can trigger short-term sell-offs and, in some cases, partial reversals as information is processed over subsequent trading days. If such dynamics occur systematically, it speaks to the efficiency with which markets incorporate ESG-related information and whether short-run prices overshoot fundamental value.

Taken together, these arguments motivate a systematic analysis of how equity markets respond to ESG controversies. This thesis examines abnormal stock returns around controversy announcements and investigates the heterogeneity of the market reaction across ESG pillars and firm characteristics that may cushion or amplify it.

Although ESG controversies are attracting more attention from investors and regulators, there are still gaps in the understanding of how equity markets price these events.

First, much of the literature on the relationship between ESG and financial performance focuses on aggregate ESG scores and long-term performance associations rather than isolating the valuation impact of specific negative incidents. This focus is partly driven by the widespread availability of ratings. However, ESG scores are imperfect proxies for event-specific information and differ significantly across providers (Friede et al., 2015; Berg et al., 2022). Therefore, evidence based on overall scores does not directly address how markets respond when specific ESG misconduct becomes public.

Second, the evidence on the severity and duration of market reactions to controversies is mixed. While many event studies document negative abnormal returns around announcement dates, the estimated effect sizes and their post-event dynamics vary across samples, event definitions, and window choices (Krüger, 2015; Capelle-Blancard and Petit, 2019; De Winne and Petkeviciute, 2022). This lack of convergence limits inference on whether price effects are primarily immediate, whether they continue as information unfolds, or whether they partially reverse.

Third, the literature provides limited integrated evidence on heterogeneity across ESG pillars and on the role of firm-level moderators. Existing studies often examine one pillar (e.g., environmental incidents) or one conditioning factor (e.g., attention or visibility), which reduces comparability and makes it difficult to identify the conditions under which controversies are most value-destructive (Aouadi and Marsat, 2018; Cui and Docherty, 2020). In particular, moderators such as firm size, profitability, leverage, and pre-event ESG reputation are not consistently incorporated within a unified empirical design.

This thesis addresses these gaps by (i) focusing explicitly on identifiable ESG controversy events, (ii) estimating short- and medium-horizon abnormal returns within a single coherent event-study framework, and (iii) testing heterogeneity by ESG pillar and firm characteristics to clarify when and why shareholder value is most affected.

This thesis contributes to the ESG-controversy event-study literature in three ways. First, it uses a manually validated, multi-country sample of initial disclosure dates from Refinitiv ESG Controversies, reducing timing noise that can bias short-horizon inference. Second, it compares Environmental, Social, and Governance incidents within one unified empirical design and consistent windows, improving comparability across pillars. Third, it tests whether pre-event ESG reputation acts as a buffer or a credibility-based backlash once firm fundamentals are controlled for, helping reconcile competing interpretations in the literature.

The primary objective of this thesis is to quantify how equity markets react to ESG controversy disclosures and to explain cross-sectional variation in these reactions. The empirical strategy relies on the event-study methodology, which isolates the stock price response to new information by comparing realized returns to model-implied expected returns (Fama et al., 1969; Brown and Warner, 1985; MacKinlay, 1997). Consistent with best practice in short-horizon event studies, the analysis uses narrow windows to reduce confounding and to improve attribution of abnormal performance to the controversy event (Kothari and Warner, 2006).

Building on this design, the thesis addresses the following research questions:

(1) **Average effect:** Do ESG controversies generate statistically and economically significant negative abnormal returns around disclosure?
(2) **Pillar heterogeneity:** Does the market reaction differ across Environmental, Social, and Governance controversies?
(3) **Moderation:** Do firm characteristics, in particular pre-event ESG reputation, firm size, profitability, and leverage moderate the magnitude of the abnormal return?
(4) **Short-run dynamics:** Do returns exhibit persistence, drift, or partial reversal when extending the analysis beyond the immediate announcement window?

Together, these questions allow the thesis to assess whether ESG controversies constitute financially material information shocks, whether markets differentiate across controversy types, and which firm attributes cushion or amplify the valuation impact.

This thesis contributes to both theory and practice. From a theoretical perspective, ESG controversies provide a natural test of semi-strong market efficiency. If controversy disclosures contain value-relevant public information, stock prices should incorporate it quickly Fama (1970). Whether abnormal returns are concentrated around the announcement date or unfold over subsequent days helps assess how investors process ESG-related news, including the role of attention, salience, and potential short-run mispricing.

The analysis also speaks to signaling and credibility. A controversy can act as a negative signal about operational quality, internal controls, or governance, and may weaken the credibility of previously observed ESG signals (including ratings). Building on evidence that ESG scores relate to both scandal risk and market reactions, Sun et al. (2024) motivate testing whether a stronger pre-event ESG reputation buffers the market penalty or, alternatively, increases it through heightened expectations and disappointment.

Practically, estimating price responses to ESG controversies supports decision-making by multiple stakeholders. Investors can use the results to improve ESG risk management and event-driven monitoring. Firms can benchmark the potential valuation impact of controversy exposure and strengthen governance and ESG controls accordingly. Regulators and standard setters can draw on the evidence when assessing the financial materiality of ESG incidents and the design of disclosure regimes.

In previewing the main results, we find that governance controversies erase approximately 2.5 times more shareholder value than environmental or social incidents within the three-day period surrounding disclosure. Furthermore, firms with robust pre-event ESG reputations lose roughly one-quarter of what low-ESG firms lose.

Chapter 2 builds the conceptual bridge from ESG controversies to stock-price reactions. It synthesizes the mechanisms that make controversies value-relevant and uses this logic to derive the hypotheses that structure the empirical tests.

Chapter 3 operationalizes these predictions. It motivates the event-study approach, defines abnormal returns and cumulative abnormal returns, documents the event and market-data sources, and explains the estimation and event windows chosen to isolate disclosure effects. The chapter also specifies the inferential strategy and robustness procedures for non-normal CAR distributions and dependence when events cluster in calendar time.

Chapter 4 implements the design on the sample. It reports sample composition and descriptive statistics, tests the hypotheses in sequence (average reaction, pillar differences, and cross-sectional heterogeneity linked to ESG reputation and firm characteristics), and stress-tests the conclusions under alternative specifications and outlier treatments.

Chapter 5 interprets what the empirical patterns imply for theory and prior evidence. It evaluates which economic channels align with the timing and magnitude of abnormal returns and clarifies how pillar heterogeneity and reputational capital shape the valuation impact.

Chapter 6 answers the research question directly and consolidates the thesis contribution. It summarizes what the evidence implies about the financial materiality of ESG controversies, states the main takeaways, and outlines limitations and directions for future research.

CHAPTER 2

Literature Review and Conceptual Framework

The literature generally documents negative average market reactions to ESG controversies, but the results vary across event definitions, timing precision, and moderator choices. This chapter therefore does two things. First, it synthesizes the main theoretical mechanisms and empirical debates. Second, it explains the specific design choices and hypotheses tested in this thesis.

2.1. Introduction and Concept Definition

ESG controversies matter in financial markets because they can change investors' expectations about a company's future cash flows and risk. More broadly, ESG-related information can affect equity valuation by shaping how long-term fundamentals and discount rates are perceived. This is especially true when the information is financially significant (Khan et al., 2016; Giese et al., 2019; Oikonomou et al., 2012). In semi-strong efficient markets, publicly available information that affects firm value is quickly reflected in stock prices (Fama, 1970). Event-study methods therefore treat controversy disclosures as information events and test whether stock prices adjust systematically (Fama et al., 1969; MacKinlay, 1997; Kothari and Warner, 2006).

For the purposes of this thesis, an *ESG controversy* is defined as a negative ESG-related event, allegation, or incident that becomes publicly known, constitutes new information, and causes reputational damage. Examples include environmental accidents, labor or customer-related misconduct, and governance failures such as corruption or fraud. This definition is consistent with prior studies that treat controversies as negative information shocks and document market penalties around disclosure dates (Krüger, 2015; Capelle-Blancard and Petit, 2019; De Winne and Petkeviciute, 2022; Li et al., 2019).

Greenwashing refers to a mismatch between a firm's ESG-related communication or reputation and its underlying practices. Investors may rely on ESG narratives or ratings, but later controversies can reveal that the reported ESG quality was overstated (Yu et al., 2020; Huang et al., 2024a). In this sense, greenwashing is a credibility problem: earlier ESG signals lose informational value once they are contradicted by visible misconduct (Spence, 1973; Connelly et al., 2011; Suchman, 1995).

A *market reaction* is captured by abnormal returns around the controversy disclosure date. Abnormal returns isolate the portion of the price change that cannot be explained by broad market movements or systematic risk factors. Event studies measure price reactions to new information by comparing realized returns with model-implied expected returns. Definitions of abnormal returns and cumulative abnormal returns used in this thesis are provided in Chapter 3 (Brown and Warner, 1985; MacKinlay, 1997; Kothari and Warner, 2006). Following this

framework, the thesis evaluates market reactions across short-, medium-, and longer-horizon windows.

2.2. Economic Mechanisms: The "Why"

ESG controversies can affect firm value by changing expected cash flows, the required rate of return, or both. From a valuation perspective, controversies matter most when they reveal financially significant downside risks or constraints that were not fully priced before the announcement (Khan et al., 2016; Giese et al., 2019). The channels below motivate testable predictions about cross-sectional differences in abnormal returns. While a clean decomposition into cash flow versus discount-rate revisions is not feasible with daily equity returns alone, the empirical analysis uses observable proxies such as pillar type, ESG reputation, and firm characteristics to assess which channel is more likely to dominate (Kothari and Warner, 2006).

2.2.1. Information and Cash-Flow Revision

A controversy can reveal information about direct cash flow effects, such as fines, litigation, remediation costs, and compliance investments, as well as indirect effects such as lost demand, canceled contracts, and supply chain disruptions. Once this information becomes public, investors revise their cash flow expectations, which should be reflected in the stock price around the announcement date (Kothari and Warner, 2006). Studies of corporate misconduct document substantial legal and reputational costs following negative announcements, which supports the cash flow revision channel (Karpoff et al., 2008). Whether the effect is limited to the announcement or continues beyond it depends on how much uncertainty the news resolves and how severe the incident is (Khan et al., 2016).

A complication, however, is that the severity and financial materiality of ESG incidents are rarely clear in advance. What initially looks like a minor allegation can turn into a major legal case, while a highly visible controversy can fade with limited direct costs. This creates an identification problem for the cash flow channel, because announcement-window CARs may reflect investors' expectations about how serious the incident will become, rather than a clean revision in discounted cash flows.

2.2.2. Reputation, Trust, and the Discount-Rate Channel

Controversies can also damage reputation and stakeholder trust beyond any direct financial costs. Firm value may fall when customers, employees, regulators, or investors reassess the firm more negatively. This can raise perceived risk and the required rate of return, even if near-term cash flows do not change immediately (Oikonomou et al., 2012; Giese et al., 2019). This channel is consistent with research that views corporate responsibility and social capital as forms of reputational insurance and trust that shape firm performance during periods of stress (Godfrey, 2005; Minor and Morgan, 2011; Lins et al., 2017). Importantly, the reputational effect may be stronger when a firm has previously projected a strong ESG image, because a controversy can be interpreted as a credibility shock rather than an isolated operational failure (Sun et al., 2024).

This creates a clear tension with the insurance view of reputational capital. The same pre-event ESG reputation that may soften the market reaction under one interpretation may intensify it under a credibility-correction view. In other words, these two mechanisms make opposite predictions, so identifying which one dominates is not just an empirical issue but also a meaningful theoretical test.

2.2.3. Agency and Governance Signal

Controversies may also signal weaknesses in internal controls and governance. From an agency perspective, governance failures raise concerns about managerial incentives, oversight quality, reporting reliability, and the likelihood of future misconduct (Jensen and Meckling, 1976; Karpoff et al., 2008). Such signals can increase uncertainty and lead to stronger price reactions, especially when investors interpret the event as evidence of broader systemic problems rather than a one-off failure. Event studies also face a practical complication here, because volatility often rises around information shocks, which can affect inference and makes robust testing methods especially important (Brown and Warner, 1985; Kolari and Pynnonen, 2010).

Taken together, these mechanisms imply that market reactions should be heterogeneous. Penalties may vary by controversy type and financial significance (Khan et al., 2016). They may also vary by firm characteristics. A strong pre-event ESG reputation or stronger governance may soften the market reaction, but it may also worsen it if the controversy violates especially high prior expectations (Sun et al., 2024; Nirino et al., 2021; Elamer and Boulhaga, 2024).

2.3. Empirical Evidence: The "What"

The empirical literature generally finds that negative ESG-related incidents are associated with negative abnormal stock performance around the time of the announcement. However, this apparent consensus hides substantial variation. Estimated effects range from economically small to double-digit losses, post-event dynamics range from quick reversals to persistent negative drift, and the conditions under which controversies are most value-destructive remain disputed. The discussion below therefore reviews the evidence critically by highlighting not only what the literature has established, but also where findings conflict, which methodological differences may explain those conflicts, and which questions remain unresolved (Krüger, 2015; Capelle-Blancard and Petit, 2019; Mu and Shi, 2023; Sun et al., 2024; De Winne and Petkeviciute, 2022). Beyond short-run price effects, several studies also examine how controversies relate to broader performance and risk outcomes. These studies point to governance arrangements and ESG practices as potential buffers (Li et al., 2019; Nirino et al., 2021; Elamer and Boulhaga, 2024).

2.3.1. Direction and Magnitude of Reactions

Most event-study findings point to a negative market response to negative ESG news within short windows around disclosure (Krüger, 2015; Capelle-Blancard and Petit, 2019; Mu and Shi, 2023; Sun et al., 2024). This pattern suggests that controversies reveal bad news about future costs, operations, or governance quality. However, effect sizes differ widely across studies, from relatively small CARs in broad samples that include minor allegations to much larger losses in studies focused on severe or highly visible incidents. This variation is unlikely to be random noise. Instead, it reflects systematic differences in event selection, sample composition, and research design that make direct comparison across studies difficult (Kothari and Warner, 2006). A key unresolved question is whether these differences reflect real variation in how markets respond to different controversies, or whether they mainly reflect differences in how studies define and select events. This thesis addresses that issue by focusing on manually validated, high-visibility events and by reporting results across multiple windows within one consistent framework.

2.3.2. Timing: Immediate Versus Lasting Effects

A central disagreement in the literature concerns whether market penalties are confined to the immediate disclosure period or continue into later trading days. Methodologically, this issue is closely tied to event-window choice and to assumptions about how quickly information is incorporated into prices (Kothari and Warner, 2006). Short-horizon designs are generally considered more reliable when event timing is precise. By contrast, longer windows make inference more difficult because confounding news and post-event volatility become more likely (Brown and Warner, 1985; MacKinlay, 1997). For this reason, many studies report multiple windows to distinguish the immediate price adjustment from the immediate post-event response (Kothari and Warner, 2006).

A deeper tension in this literature is whether the absence of post-announcement reversal should be read as evidence of efficient pricing or as evidence of continued information arrival. If a controversy unfolds over several days through follow-up news, regulatory responses, or analyst revisions, then negative CARs in medium-horizon windows may partly reflect new information rather than the slow digestion of the original disclosure. This ambiguity limits what can be inferred from longer windows and is one reason this thesis reports immediate and subsequent horizons separately.

2.3.3. Heterogeneity by Pillar, Severity, and Firm Characteristics

The literature also suggests that reactions vary across controversy types and firm contexts. These differences may reflect variation in the perceived seriousness of different ESG dimensions, in their implications for internal controls, and in how clearly incidents signal financial consequences (Krüger, 2015; Aouadi and Marsat, 2018; Sun et al., 2024). Firm characteristics matter as well. Size, profitability, and governance structure can shape both the underlying financial impact of a controversy and the ease with which abnormal performance can be detected (Kothari and Warner, 2006). Additional evidence links controversies to broader performance and risk-taking outcomes, especially in financial institutions (Li et al., 2019; Brighi et al., 2023).

Even with this growing evidence, the literature still lacks a unified framework that tests pillar differences and firm-level moderators within one consistent design. Studies that focus on only one pillar cannot show whether, for example, governance failures are inherently more value-destructive or whether their stronger effects in some papers reflect sample composition. Likewise, studies that examine only one moderator cannot show whether that moderator matters on its own or simply proxies for size, profitability, or ESG reputation. This fragmentation of the evidence base is the main methodological gap that motivates the integrated design of this thesis.

2.3.4. ESG Ratings as Buffer Versus Backlash: The Greenwashing Link

A prominent debate asks whether ESG ratings and reputational capital reduce or intensify market penalties.

Buffer logic: A strong ESG reputation or stronger ESG practices may reduce the damage if stakeholders see an incident as less severe or if reputational capital provides resilience (Nirino et al., 2021; Elamer and Boulhaga, 2024).

Backlash logic grounded in greenwashing: A greenwashing-based backlash offers one explanation for why high-ESG firms may be punished more strongly after a controversy. If a firm's ESG signals are not fully supported by its underlying practices, a controversy may be interpreted not only as an isolated incident but also as a credibility shock (Yu et al., 2020; Huang et al., 2024a). In that case, investors update in two ways at the same time. First, they price the direct costs of the controversy itself. Second, they revise downward their beliefs about the reliability of the firm's earlier ESG signals. The combination of incident costs and credibility loss is what separates a greenwashing-driven backlash from a simple negative surprise.

This interpretation is consistent with earlier evidence. Yu et al. (2020) show that ESG disclosures can diverge from underlying corporate practices, while Huang et al. (2024a) document ESG-washing behavior in banking.

Overall, the mixed evidence motivates testing whether pre-event ESG reputation moderates market reactions and whether the direction of that moderation is more consistent with a buffering mechanism or with a credibility-based backlash.

2.4. Research Gap

Despite the growing body of evidence on ESG controversies, three issues remain unresolved.

First, the literature provides mixed evidence on whether a strong pre-event ESG reputation reduces or increases controversy-related value loss. Reputational-capital arguments predict a buffering effect, whereas greenwashing-based arguments predict a backlash when controversies reveal a gap between ESG signals and underlying practices (Yu et al., 2020; Huang et al., 2024b; Sun et al., 2024). Existing empirical findings do not clearly establish which mechanism dominates across settings and incident types.

Second, few studies compare short- and medium-term post-event effects within one consistent framework. This matters because short windows minimize confounding noise, while medium windows reveal whether the effect persists (MacKinlay, 1997; Kothari and Warner, 2006). Reporting multiple windows within a single design therefore helps distinguish the initial price adjustment from short-term follow-up effects.

Third, it remains unclear which ESG pillar produces the strongest negative reaction in comparable settings. While cross-category comparisons do exist, the results vary with region, event definitions, and modeling choices (Brighi et al., 2023; Krüger, 2015; Capelle-Blancard and Petit, 2019). A systematic design that classifies controversies by pillar and tests differences within the same empirical framework can therefore improve comparability.

These gaps motivate the central research question: *How do equity markets react to ESG controversies across different horizons, and how do pillar type, ESG reputation, and firm characteristics shape the magnitude of abnormal returns?*

2.5. Hypotheses

Building on the mechanisms discussed above, the empirical analysis tests whether ESG controversies lead to abnormal stock price reactions, whether reactions differ across ESG pillars, and whether firm-level resilience factors reduce the size of the penalty. Market reactions are measured using CAR and CAAR across the short, medium, and long event windows defined in Chapter 3 (Brown and Warner, 1985; Kothari and Warner, 2006).

H1: Baseline Effect

ESG controversies generate negative abnormal returns around the announcement date.

Rationale: A controversy is negative information. It can trigger revisions in expected cash flows through direct and indirect costs, and it can also increase perceived risk through reputational damage. Both channels imply lower firm value (Kothari and Warner, 2006).

H2: Pillar Heterogeneity

Governance controversies lead to more negative abnormal returns than environmental or social controversies.

Rationale: Governance incidents more directly signal control failures and agency problems. They can therefore raise stronger concerns about reporting integrity, oversight quality, and the likelihood of future misconduct. This channel predicts a more negative CAR or CAAR for governance-related events (Kothari and Warner, 2006; Karpoff et al., 2008).

H3: Moderators of Controversy-Related Value Loss

The magnitude of abnormal returns following ESG controversies depends on firm-level resilience and credibility.

Rationale: The same controversy can create different expected costs and risks depending on the firm's reputation, size, and operating strength before the event. Using the empirical design from Chapter 4.4, this thesis tests moderation by comparing subgroups based on these characteristics.

H3a: ESG reputation (buffer vs. backlash): Firms with stronger pre-event ESG reputation experience less negative abnormal returns after ESG controversies. *Rationale:* Reputational capital may work as a form of insurance, reducing stakeholder and investor penalties (Godfrey, 2005; Minor and Morgan, 2011; Lins et al., 2017). A competing view in the literature is backlash in cases of perceived greenwashing. This moderator is therefore tested explicitly (Yu et al., 2020; Huang et al., 2024b).

H3b: Firm size Higher Market Cap: Larger firms experience less negative abnormal returns than smaller firms following ESG controversies. *Rationale:* Larger firms usually have greater financial and organizational capacity to absorb shocks and respond credibly. They also tend to receive more investor attention, which can affect information processing and trading dynamics (Cui and Docherty, 2020).

H3c: Profitability: More profitable firms experience less negative abnormal returns following ESG controversies. *Rationale:* Higher profitability signals operational quality and financial flexibility, which may reduce concerns about distress and remediation costs in the immediate aftermath of a controversy.

H3d: Leverage: Financial leverage does not systematically moderate short-window abnormal returns following ESG controversies. *Rationale:* Theoretical predictions for leverage are ambiguous in the short run. Higher leverage may amplify the reaction through financial distress risk, but it may also reduce agency concerns through stronger monitoring by creditors. This justifies testing the moderator empirically without imposing a strong directional prior.

Together, H1 through H3 provide a coherent structure that mirrors the empirical organization in Chapter 4, namely the baseline effect, pillar heterogeneity, and moderator tests.

CHAPTER 3

Methodology

3.1. Research Design and Approach

This study uses a quantitative, deductive research design. The study focuses on an event study to examine how stock markets respond to events involving ESG controversies. The hypotheses are grounded in three theoretical frameworks. First, semi-strong market efficiency suggests that stock prices adjust quickly to public information (Fama, 1970). Second, stakeholder and legitimacy theory imply that ESG failures can trigger reputational damage and stakeholder withdrawal (Suchman, 1995). Third, agency theory views governance controversies as signals of control failure and higher agency costs (Jensen and Meckling, 1976). The event-study method isolates the effect of a specific event on stock prices by measuring abnormal returns. Abnormal returns are the difference between a stock's actual return and its expected normal return. This method, pioneered by Fama et al. (1969), is standard in financial economics for studying how quickly markets absorb new information. It provides a direct measure of the change in shareholder wealth caused by an event under the efficient market hypothesis, which assumes that new information is quickly reflected in prices. By focusing on short-term price changes around controversy announcements, I capture the market's immediate evaluation of these incidents.

This approach is hypothesis-driven and follows best practices (Brown and Warner, 1985; MacKinlay, 1997) for event studies. I focus on carefully identifying event dates and clearly defining the estimation period for normal performance. Short-horizon event studies are known to provide reliable results (Kothari and Warner, 2006). In contrast, long-term analyses often encounter interfering factors. Therefore, focusing on the immediate period surrounding each event yields more robust inferences about its impact. Overall, this event-study design enables rigorous testing of the cause-effect link (ESG controversy $\rightarrow$ stock return impact) within a deductive research paradigm.

3.2. Data and Sample

3.2.1. Data Source

I obtain ESG controversy events from the Refinitiv (LSEG) ESG Controversies database. The database records firm-specific, ESG-related incidents and assigns each one to a primary ESG pillar. Refinitiv provides structured event identifiers and associated disclosure information, which enables me to construct event dates for an event-study design. Importantly, the database serves as the initial set of potential incidents. The final event sample is constructed using explicit inclusion criteria and a manual validation step, which are described below.

3.2.2. Sample Selection

The sample covers listed firms from global equity markets between 2015 and 2024. Starting from the Refinitiv ESG Controversies universe, I construct an event-study sample using a two-stage procedure: (i) mechanical eligibility filters and (ii) manual validation of event timing and relevance.

Stage 1: Eligibility filters. An observation enters the candidate set if it satisfies the following criteria:

(1) The firm is publicly listed and has daily return data available for the full estimation window and all event windows.
(2) The controversy has an identifiable initial public disclosure date that can be mapped to a trading day (day 0).
(3) Events are excluded if return data are missing or if trading is too infrequent to compute reliable daily returns.
(4) To limit statistical dependence, overlapping events for the same firm within the event window are excluded.

Stage 2: Manual validation (event timing and salience). Because controversy disclosures can unfold over multiple days and database timestamps may reflect updates rather than first disclosure, I manually validate each candidate event. Specifically, I verify that the assigned day-0 date corresponds to the first major public disclosure of the incident (rather than follow-up reporting) and that the incident content matches the ESG pillar classification. When disclosures span multiple days, I retain the earliest major disclosure as the event date.

After applying these steps, the final sample comprises 114 controversy events. Summary descriptives are reported in Table 4.1.

3.2.3. Stock Returns and Market Data

Daily stock returns are obtained from Refinitiv. Expected returns are estimated using a home-market benchmark approach: each firm is matched to a broad equity market index that corresponds to its primary listing venue, inferred from the Refinitiv RIC exchange extension. This aligns the market factor with the trading environment in which the stock is priced and mitigates currency and market-integration distortions that can arise when applying a single global index to all firms. Appendix A reports the benchmark mapping used in the empirical analysis.

3.3. Operationalization and Variables

In this section, I define the key variables and explain how I constructed them for the analysis.

Table 3.1. List of equations used in the thesis

Eq.	Equation
(3.1)	$AR_{i,t} = R_{i,t} - \widehat{E}(R_{i,t})$
(3.2)	$\widehat{E}(R_{i,t}) = \widehat{\alpha}_i + \widehat{\beta}_i R_{m,t}$
(3.3)	$CAR_i(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} AR_{i,t}$
(3.4)	$\text{Leverage}_{i,t-1} = \frac{\text{Debt}_{i,t-1}}{\text{Debt}_{i,t-1} + \text{Equity}_{i,t-1}}$
(3.5)	$t = \frac{\overline{CAR}}{s/\sqrt{N}}$
(3.6)	$CAR_i = \gamma_0 + \gamma_1 \text{Large}_i + \gamma_2 \text{HighProfit}_i + \varepsilon_i$

3.3.1. Normal and Abnormal Returns

The main dependent variable in the event study is the abnormal return (AR). For firm i on day t, I define

$$AR_{i,t} = R_{i,t} - \widehat{E}(R_{i,t}) \tag{3.1}$$

where $R_{i,t}$ is the stock's realized return and $\widehat{E}(R_{i,t})$ is the expected return on that day absent the event. To estimate expected returns, I use the **market model**. This is a standard approach where a stock's normal performance is predicted from its historical relationship with the market. Specifically, I run a firm-specific OLS regression of $R_{i,t}$ on the market index return $R_{m,t}$ over a pre-event estimation window. This yields an intercept (α_i) and slope (β_i). The expected return on day t is then

$$\widehat{E}(R_{i,t}) = \widehat{\alpha}_i + \widehat{\beta}_i R_{m,t} \tag{3.2}$$

The abnormal return is the difference between the actual and predicted returns. This market-adjusted AR captures the part of the stock move linked to firm-specific news. The market model is chosen over alternative expected-return benchmarks based on three main considerations. First, unlike the constant-mean return model, the market model accounts for systematic, market-wide movements on each event day. This is especially important in a multi-country sample because market conditions vary across countries and over time. In this study, expected returns are estimated using a country-specific benchmark index for each stock so that abnormal returns can be measured relative to the firm's domestic market. This improves the precision of abnormal return estimates and reduces noise in cross-sectional inference (Brown and Warner, 1985; MacKinlay, 1997).

Second, the market model avoids the need to estimate a risk-free rate for each firm-country combination across 24 markets and multiple currencies, unlike the CAPM. Using a single global

risk-free rate would create distortions, and using country-specific risk-free rates would add complexity and measurement uncertainty without materially improving short-term inference (Brown and Warner, 1985; MacKinlay, 1997).

Third, the market model is more suitable for a global event study than multi-factor models, such as the Fama-French three-factor or four-factor models, because daily factor data are not consistently available for all markets in the sample. As Brown and Warner (1985) show, in short-horizon event studies, the gains from adding further factors are usually limited because the market factor already captures most of the systematic variation in daily stock returns.

3.3.2. Cumulative Abnormal Return (CAR)

Although I calculate abnormal returns daily, the primary outcome measures are the **cumulative abnormal returns** (CAR). For event i and an event window from day τ_1 to τ_2, I define

$$CAR_i(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} AR_{i,t} \tag{3.3}$$

For example, $CAR_i(-1, +1)$ is the sum of abnormal returns on the day before the event, the event day, and the day after. CAR measures the total abnormal return associated with the event over several days, which is useful when the price reaction is spread over more than one trading day. In my analysis, I compute CARs for several event windows (defined in Section 3.4) to compare immediate reactions with slightly longer reactions. I use one CAR per event as the outcome variable capturing the controversy's impact on shareholder value. A negative CAR indicates value loss linked to the controversy, while a positive CAR can occur in rare cases if the disclosure coincides with offsetting favorable information.

Averaging $CAR_i(\tau_1, \tau_2)$ across events yields the cumulative average abnormal return (CAAR).

3.3.3. ESG Pillar Classification

Each event is assigned to its main ESG dimension: Environmental (E), Social (S), or Governance (G). This classification is based on the Refinitiv system, with manual checks performed when necessary. If an event relates to more than one ESG area, it is assigned to the most appropriate pillar. This prevents the same event from being counted in more than one category. For instance, an oil spill would be classified as environmental, a labor rights violation as social, and a corporate fraud case as governance. Using one main pillar per incident keeps the categories mutually exclusive for the analysis.

I create three pillar dummy variables: D_i^E equals 1 if event i is Environmental (0 otherwise), D_i^S equals 1 if event i is Social (0 otherwise), and D_i^G equals 1 if event i is Governance (0 otherwise).

In some analyses, I exclude one category (e.g., I use governance as the reference group in regressions) to avoid perfect multicollinearity. In descriptive comparisons, however, I consider all three groups. I expect to see differences across pillars (Hypothesis 2).

3.3.4. Firm Characteristics and Dummy Variables

For each moderator, I created a binary indicator based on the sample median, following prior event studies (Krüger, 2015; Capelle-Blancard and Petit, 2019). This approach has two advantages. First, it does not impose a linear relationship between the moderator and abnormal returns, which may be too restrictive in a sample of 114 events. Second, it yields subgroup comparisons that are easy to interpret economically, such as large versus small firms. Its main limitation is lower statistical power than continuous specifications. I therefore complement the subgroup analysis with cross-sectional OLS regressions that treat the moderators as continuous variables, as described in Section 3.5 and Appendix A.

- **Size Dummy:** I define $\text{Large}_i = 1$ if company i is large (above the sample median market capitalization) and $\text{Large}_i = 0$ if it is smaller (below the median). Market capitalization is taken from the fiscal year before the event. Market capitalization is the standard proxy for firm size in event-study and ESG research. It directly captures the economic scale of the firm as perceived by investors at the time of the event. Market capitalization is also consistently available across listing markets (Fama and French, 1993; Kothari and Warner, 2006). Larger firms often have greater organizational capacity to absorb reputational shocks, more diversified investor bases, and higher analyst coverage. These factors can influence how quickly and fully controversy-related information is priced (Cui and Docherty, 2020). An alternative proxy is total assets. However, total assets are based on book values and accounting conventions that vary across countries and industries. This makes market capitalization a more comparable measure in this cross-country sample.
- **Profitability Dummy:** I define $\text{HighProfit}_i = 1$ for firms with high profitability and 0 for firms with lower profitability. Profitability is measured by return on assets (ROA), calculated as $\frac{\text{Net Income}}{\text{Total Assets}}$ for the fiscal year prior to the event. Similar to the size dummy, I classify firms based on the sample median ROA: if a firm's ROA in that year is above the median, then $\text{HighProfit}_i = 1$; otherwise $\text{HighProfit}_i = 0$. ROA is widely used as a profitability proxy in event-study and corporate finance research because it measures how efficiently a firm converts its assets into earnings, and it remains comparable across firms of different sizes and capital structures (Krüger, 2015; Oikonomou et al., 2012).
- **Leverage Dummy:** I define $\text{HighLeverage}_i = 1$ if the firm's leverage is above the sample median (and 0 if below), using accounting data from the fiscal year prior to the event ($t-1$). Leverage is measured as the debt to capital ratio, defined as total debt divided by total debt plus equity, following prior research in capital structure and ESG event studies (Frank and Goyal, 2009; De Winne and Petkeviciute, 2022). This measure captures the share of firm financing that comes from debt and is bounded between zero and one, which makes it easier to compare across firms and markets.

$$\text{Leverage}_{i,t-1} = \frac{\text{Debt}_{i,t-1}}{\text{Debt}_{i,t-1} + \text{Equity}_{i,t-1}} \tag{3.4}$$

- **ESG Reputation Dummy:** I define $\text{HighESG}_i = 1$ for firms with an above-median pre-event ESG score (and 0 for firms with a below-median score). The pre-event Refinitiv ESG score is used as a proxy for reputational capital because it provides a composite and publicly observable measure of the firm's ESG standing before the controversy (Sun et al., 2024; Aouadi and Marsat, 2018). Composite ESG scores are widely used in the controversy literature because they combine multiple dimensions of corporate behaviour and therefore capture the overall credibility signal that investors and other stakeholders may rely on when interpreting negative news.

  At the same time, ESG ratings differ substantially across providers because of differences in scope, weighting, and measurement choices (Berg et al., 2022). Since this study relies on Refinitiv only, the results should be interpreted as conditional on this specific measure of ESG reputation and may not fully generalise to other rating systems.

These variables are used in the heterogeneity analyses (see Section 3.5). For example, I compare CAARs for large versus small firms, high versus low profitability firms, and so on. This shows whether the market reaction differs systematically across these firm characteristics. All key equations are summarized in Table 3.1.

3.4. Event-Study Design

Next, I explain how I implement the event study. This includes the estimation window, event windows, and how I compute abnormal returns and CARs for the sample.

3.4.1. Estimation Window

I estimate the market model parameters for each event using pre-event historical data. For short-term event studies, the typical range is 250 trading days prior to the event. I use a relatively long estimation window to obtain stable estimates of α_i and β_i. In practice, I use an estimation period of $[-250, -20]$ trading days relative to the event date (day 0). The 20-day gap between the end of the estimation window and day 0 acts as a buffer. It reduces the risk that the estimation period is contaminated by event-related price adjustments close to the disclosure date. Using approximately 230 trading days of data should provide precise parameter estimates (Sun et al., 2024; Brown and Warner, 1985).

3.4.2. Event Windows

I define several event windows around each controversy to measure the stock's reaction over different horizons. The main window is the short-term **3-day window** $[-1, +1]$. It includes the trading day before the event, the event day, and the day after. This window captures three effects: a possible leak or rumor before the event, the immediate reaction on the announcement day, and a one-day adjustment after the event (MacKinlay, 1997; Brown and Warner, 1985).

I include day -1 to account for potential timing frictions and short-horizon information diffusion prior to the recorded disclosure date. Nevertheless, I cannot fully exclude earlier leakage or gradual information incorporation before day -1, which is a general limitation when event dates are derived from news-based controversy databases.

In addition, I use a **one-week window** $[-1, +5]$ (7 trading days) and a **one-month window** $[-1, +20]$ (21 trading days). These longer windows help to see whether the effect lasts longer or whether the reaction is delayed. For example, the $[-1, +5]$ window covers about one week. It can show whether the stock keeps drifting down (or starts to recover) after the first news. The $[-1, +20]$ window covers about four weeks after the event. It captures any longer trend or reversal after the controversy. However, longer windows require caution. The further I move away from the event date, the more likely it is that other news or market movements affect the stock. This can weaken or blur the measured effect of the ESG event. I interpret results from the longest window with this caution in mind (Kothari and Warner, 2006; Brown and Warner, 1985; MacKinlay, 1997).

3.4.3. Abnormal Return Calculation

Using the market-model parameters $\hat{\alpha}_i$ and $\hat{\beta}_i$ estimated over the estimation window section 3.4.1, I compute abnormal returns for each trading day t in the event window according to eq. (3.1), using expected returns from eq. (3.2). I then cumulate these abnormal returns over the event window to obtain the event-level cumulative abnormal return, $CAR_i(\tau_1, \tau_2)$, as defined in eq. (3.3).

For the full sample of N events, I also report aggregate measures. Specifically, I compute the average abnormal return (AAR) for each event-time day t as the cross-sectional mean of $AR_{i,t}$ across events, and the cumulative average abnormal return (CAAR) for each window as the cross-sectional mean of the corresponding event-level $CAR_i(\tau_1, \tau_2)$ values. These aggregates summarize the average market reaction, while the main statistical tests are conducted on the distribution of event-level CAR values.

3.4.4. Expectations and Sign Convention

Because ESG controversies are usually bad news for a company (e.g., legal penalties, clean-up costs, reputational damage), I expect most events to create negative abnormal returns. My main hypothesis is that the average CAR in the sample will be significantly **negative**. I test this in Section 3.5. In rare cases, a controversy may also come with a positive element. For example, if a scandal leads to the removal of a clearly weak CEO, investors may react positively to that leadership change. However, these cases are exceptions. Overall, I expect ESG scandals to reduce shareholder value in the short run (Sun et al., 2024). This event-study design is intended to capture and measure this expected negative effect.

3.4.5. Multiple Events and Independence

In cross-sectional tests, event studies typically treat each event as an independent observation. However, this assumption may be violated if events overlap for the same firm or if many events occur on the same calendar date, which can induce cross-sectional dependence in abnormal returns (MacKinlay, 1997; Kolari and Pynnonen, 2010). Therefore, I exclude overlapping or repeated events for the same firm. In addition, as a robustness check for same-day event clustering, I re-estimate the mean event-level CAR using an intercept-only regression with standard errors clustered by calendar event date; the resulting clustered-by-date test statistics are reported in Appendix A.

3.5. Statistical Tests and Inference

After computing abnormal returns and CARs, I test for statistically significant differences in the average effects and examine how reactions vary across event types and firm characteristics. To assess robustness to distributional assumptions and outliers, I report both parametric and non-parametric tests.

3.5.1. Tests for Abnormal Return Significance

First, I ask whether ESG controversy events affect stock returns systematically. For each event window, I test the null hypothesis that the average cumulative abnormal return is zero $H_0 : \mathbb{E}[CAR] = 0$. I begin with a one-sample *t*-test using the event-level CAR values. The test statistic is

$$t = \frac{\overline{CAR}}{s/\sqrt{N}}. \tag{3.5}$$

where $\overline{CAR}$ is the average CAR across the N events and s is the standard deviation of CAR in the sample. Assuming the events are independent, this statistic follows a *t*-distribution with $N - 1$ degrees of freedom. I report two-tailed p-values at the 5% level (MacKinlay, 1997; Brown and Warner, 1985).

Because multiple events can share the same calendar date, I also report calendar-date clustered inference (Appendix A) to ensure that statistical significance is not driven by same-day cross-sectional dependence.

3.5.2. Non-Parametric Tests

Event-level CAR values may not follow a normal distribution and may include outliers. To avoid relying only on normality assumptions, I add non-parametric tests to my analysis.

First, I use a sign test based on the share of negative CAR values. Under the null hypothesis of no overall effect, a negative CAR should occur with probability 0.5. I test this with a two-sided binomial test that checks whether the observed share of negative CAR values differs from 0.5.

Second, I use the Wilcoxon signed-rank test, which uses both the sign and the size of CAR values. The test ranks the absolute CAR values and then compares the ranks of positive and negative observations. It checks whether the distribution is centered around zero. Together, these tests are less sensitive to outliers and do not require normality.

3.5.3. Comparing ESG Pillar Reactions

One of my main goals is to test whether market reactions differ across the three pillars of ESG. After computing event-level CAR values, I will split the events into three groups by pillar and compare the results. Since I am comparing three independent groups and do not want to assume a normal distribution, I use the Kruskal–Wallis test. It is a rank-based, nonparametric alternative to a one-way ANOVA. The null hypothesis is that the CAR distributions are the same across the three pillars (Kruskal and Wallis, 1952).

If the Kruskal–Wallis test shows a statistically significant difference, I then run pairwise comparisons using Mann–Whitney U tests for each pillar pair (E vs. S, E vs. G, and S vs. G) (Mann and Whitney, 1947). I interpret these comparisons in light of Hypothesis 2, which predicts clear differences in the size of price reactions across pillars.

3.5.4. Heterogeneity by Firm Characteristics

To test whether market reactions differ based on firm characteristics, I examine differences in size, profitability, leverage, and pre-event ESG reputation (Hypothesis 3 and its sub-hypotheses). First, I perform split-sample comparisons using median splits for each characteristic. For each dummy variable (e.g., Large_i), I compare the average CAR between the two groups. I use a two-sample *t*-test and also report a Mann–Whitney U test to avoid relying too much on normality (Mann and Whitney, 1947).

Second, I run cross-sectional OLS regressions at the event level to test several moderators at the same time. The baseline model is

$$CAR_i = \gamma_0 + \gamma_1 \text{Large}_i + \gamma_2 \text{HighProfit}_i + \varepsilon_i. \tag{3.6}$$

and I extend it by adding further dummy variables such as HighLeverage_i and HighESG_i. The coefficients γ_k measure the average difference in CAR between the relevant groups. Because the number of events is limited ($N = 114$), I keep the models simple. I limit the number of control variables to reduce the risk of overfitting (Wooldridge, 2010).

3.6. Robustness Checks

To determine if the main findings are affected by distributional issues or sample composition, robustness checks are performed on event-level CAR measures for each event window. Detailed robustness results are reported in Section 4.5.

3.6.1. Sensitivity to Outliers

I re-estimate CAARs after applying alternative treatments of extreme CAR observations. Specifically, I (i) winsorize event-level CAR values at the 1% and 5% quantiles and (ii) trim the distribution by excluding observations below the 5th percentile and above the 95th percentile. These checks determine whether the baseline CAAR estimates are driven by a small number of extreme events.

3.6.2. Temporal Stability

To examine temporal stability, I split the sample into two subperiods: pre-2020 (2015–2019) and 2020–2024. I then repeat the main CAAR analysis in each subperiod. This assesses whether the estimated market reaction is concentrated in a particular time period.

3.6.3. Robust Inference in Cross-Sectional Regressions

For cross-sectional regressions explaining heterogeneity in CAR, I report heteroskedasticity-robust standard errors (HC3) to reduce sensitivity to heteroskedastic residuals.

CHAPTER 4

Empirical Results

This chapter presents the empirical findings on stock market reactions to ESG controversy events, organized into the following sections. Section 4.1 describes the sample composition and key descriptive statistics. Section 4.2 reports the main event study results for the full sample of 114 events. Section 4.3 examines heterogeneity in the market reaction by the ESG pillar of the controversy (Environmental, Social, or Governance). Section 4.4 investigates whether firm-level characteristics (e.g., pre-event ESG performance, size, financial strength) moderate the impact. Section 4.5 presents robustness checks to validate the findings. Finally, Section 4.6 summarizes the key results and the degree of support for each hypothesis.

4.1. Sample Composition & Descriptive Statistics

The final sample comprises 114 ESG controversy events between 2015 and 2024. Events are classified by the three ESG pillars: 27 Environmental incidents (23.7%), 46 Social incidents (40.4%), and 41 Governance incidents (36.0%). These proportions are reported in Table 4.1.

Table 4.1. Descriptive statistics of ESG controversy events

Event Category	N	% of Total	Event Year Range
Environmental	27	23.7%	2015–2023
Social	46	40.4%	2015–2024
Governance	41	36.0%	2015–2024
Total	**114**	**100.0%**	**2015–2024**

4.1.1. Firm Characteristics

Table 4.2 reports descriptive statistics for key firm characteristics measured at the fiscal year-end prior to the event year ($t-1$). The sample consists predominantly of large firms: market capitalization averages 56,795 USD million (median 23,323 USD million), while total assets average 126,834 USD million (median 31,385 USD million). Leverage has a mean of 0.467 (median 0.461) and ranges from 0.013 to 0.985. Profitability, measured by ROA, averages 0.033 (median 0.030) with values between -0.465 and 0.201. Pre-event ESG scores average 63.3 (median 68.0) and range from 9.2 to 94.7.

Table 4.2. Descriptive statistics of firm characteristics (t–1)

Characteristic	N	Mean	Median	Std. Dev	Min	Max
Total Assets (USD m)	112	126,834	31,385	279,369	237	1,854,259
Leverage	105	0.467	0.461	0.235	0.013	0.985
ROA	112	0.033	0.030	0.075	–0.465	0.201
Market Cap (USD m)	112	56,795	23,323	89,606	49	512,759
ESG Score	110	63.3	68.0	21.5	9.2	94.7

Note: Total Assets and Market Cap are presented in USD millions. ROA is return on assets. ESG Score ranges from 0 to 100.

Building on the event-study methodology described in chapter 3, the next sections present the empirical results for abnormal returns across the defined event windows.

4.2. Main Event Study Results

Table 4.3 reports the cumulative average abnormal returns (CAARs) for the 114 ESG controversy events over several event windows. The market reaction is negative and statistically significant in all windows examined. For the short window $[-1, +1]$, the CAAR is -12.43% (Std. Error = 1.832), with $t = -6.785$ ($p < 0.001$). Over the medium window $[-1, +5]$, the CAAR is -19.29% ($t = -7.083$, $p < 0.001$). For the long window $[-1, +20]$, the CAAR is -19.16% ($t = -7.091$, $p < 0.001$), indicating no evidence of reversal within the 20-trading-day horizon. Because longer windows may include follow-up disclosures and other confounding firm news, the $[-1, +20]$ estimate is interpreted as the cumulative short-run response rather than as a clean long-run valuation effect. Non-parametric tests yield similar inferences: as reported in Table 4.3, Wilcoxon signed-rank tests also confirm that the median abnormal return is significantly below zero in each window.

Table 4.3. Cumulative average abnormal returns (CAAR) by event window

Window	N	CAAR (%)	Std. Error	t-stat	p-value
Short [–1,+1]	114	–12.43	1.8320	–6.785	<0.001
Medium [–1,+5]	114	–19.29	2.7230	–7.083	<0.001
Long [–1,+20]	114	–19.16	2.7020	–7.091	<0.001

Notes. CAAR is reported in percent. All results are significant at the $p < 0.001$ level (two-sided t-tests and Wilcoxon signed-rank tests confirm significance). Event windows: Short [–1,+1], Medium [–1,+5], Long [–1,+20].

To illustrate the economic magnitude, applying the short-window CAAR of -12.43% to a firm with the sample median market capitalization (about 23,323 USD million) implies a market value loss of roughly *USD 2.9 billion* over the three-day event window. Over the long $[-1, +20]$ window, the corresponding implied loss is approximately *USD 4.5 billion* for a median-sized firm. Overall, these results indicate a pronounced and persistent negative stock price reaction

to ESG controversy events. The next section examines whether the magnitude of this reaction differs across ESG categories (H2).

To address concerns about cross-sectional dependence and event-induced variance, Appendix A.2 presents adjusted estimates. The main conclusions remain unchanged.

4.3. Heterogeneity by ESG Pillar

I next analyze whether the stock market impact of a controversy varies by the ESG pillar category of the incident (Environmental, Social, or Governance), as posited in Hypothesis 2. Table 4.5 reports the average abnormal returns for events in each category, along with tests of differences across categories.

All three subcategories (E, S, and G controversies) show significantly negative abnormal returns on average. For Environmental-related events ($N = 27$), the mean three-day CAR $[-1, +1]$ is -7.88% ($t = -4.10$, $p < 0.001$). Social-issue events ($N = 46$) have a mean $[-1, +1]$ CAR of -8.00% ($t = -5.90$, $p < 0.001$). Governance-related events ($N = 41$) show a far larger impact, with a mean $[-1, +1]$ CAR of -20.41% ($t = -4.60$, $p < 0.001$). In fact, the average three-day CAR for Governance controversies is roughly 2.5 times more negative than that for Environmental or Social events (which are around -8% each).

Statistical tests indicate heterogeneity across these groups. For example, a Kruskal–Wallis test comparing the CARs by category over the five-day window $[-1, +5]$ is significant ($p = 0.037$), indicating that at least one category differs from the others. Post-hoc pairwise comparisons (Holm-adjusted) suggest that the Governance vs. Environmental difference is suggestive ($p \approx 0.07$) and the Governance vs. Social difference is similar in magnitude, with suggestive evidence in the Holm-adjusted comparison ($p \approx 0.075$).

Table 4.4. Cross-sectional regression of CAR (medium window)

Variable	Coefficient	Std. Err (HC3)	t-stat	p-value	Sig
const	-0.1061	0.0272	-3.9	0.0001	***
Pillar_Social	-0.0503	0.0422	-1.191	0.2336	
Pillar_Governance	-0.1848	0.0683	-2.705	0.0068	***
R^2	0.0690				
Adj. R^2	0.0522				
N	114.0000				

Notes. OLS with heteroskedasticity-robust standard errors (HC3).

For the medium window $[-1, +5]$, Environmental events have an average CAR of -10.61% (significant at $p < 0.001$), Social events average -15.64% ($p < 0.001$), and Governance events average -29.09% ($p < 0.001$). By the 20-day window $[-1, +20]$, Governance events still show the most severe impact (mean CAR $\approx -27.37\%$) compared to Social (-17.72%) and Environmental (-9.14%) events. Overall, these results are consistent with H2: Governance-related controversies are associated with economically larger stock price declines than Environmental

or Social controversies, although the Holm-adjusted pairwise differences provide only marginal (suggestive) statistical support.

Table 4.5. CAAR by ESG pillar and event window

Pillar	Short	Medium	Long	KW p	G vs E/S p
Environmental	-7.88	-10.61	-9.14	0.037	0.070/0.075
Governance	-20.41	-29.09	-27.37	0.037	–
Social	-8.00	-15.64	-17.72	0.037	0.075

Notes: CAAR in percent (Short: [–1,+1], Medium: [–1,+5], Long: [–1,+20]). KW p = Kruskal–Wallis test (medium window). G vs E/S p = Governance vs Environmental/Social (Holm-adjusted p-values). All Governance differences are significant at $p < 0.10$.

4.4. Moderation Analysis

I now examine factors that might moderate the impact of ESG controversies on stock returns, as hypothesized in H3 and its sub-components. Specifically, I consider three firm-specific characteristics as potential moderators: (a) the firm's pre-event ESG reputation, (b) firm size, and (c) financial health (using leverage and profitability as indicators).

4.4.1. ESG Reputation as a Moderator (H3a)

Sub-hypothesis H3*a* posits that a firm's prior ESG reputation mitigates the severity of the market reaction to a controversy. I test this by comparing firms in the top and bottom ESG score quartiles (High: $\geq$75th percentile; Low: $\leq$25th percentile), reported in Table 4.6.

Consistent with a reputational buffering pattern, high-ESG firms exhibit substantially less negative abnormal performance than low-ESG firms across all windows. In the short window $[-1, +1]$, the mean CAR is -6.53% for high-ESG firms versus -22.88% for low-ESG firms (difference: 16.35 pp; $p = 0.01$). In the medium window $[-1, +5]$, the gap remains economically large (13.20 pp) but is only marginally significant ($p = 0.07$). In the long window $[-1, +20]$, the difference widens again to 18.15 pp ($p = 0.01$).

To mitigate omitted-variable concerns, Appendix A reports multivariate regressions of event-level CARs on ESG reputation while controlling for firm size, profitability (ROA), leverage, and pillar composition. ESG reputation remains positively associated with CARs in the short window and is also positive in the long window, consistent with the bivariate split-sample evidence. Multicollinearity is low (Appendix Table A.4).

4.4.2. Firm Size as a Moderator (H3b)

Hypothesis H3*b* conjectures that larger firms experience less severe stock price reactions to ESG controversies than smaller firms. To test this, I split the sample by firm size at the median market capitalization, yielding two groups of roughly equal size (large-cap vs. small-cap firms). The results are presented in Table 4.7.

Table 4.6. ESG reputation and market reaction (high vs. low ESG)

ESG Group	N	Mean ESG	Short	Medium	Long
High (≥75th)	28	86.1	-6.53	-10.45	-7.70
Low (≤25th)	28	31.6	-22.88	-23.64	-25.85
Difference			16.35	13.20	18.15
t-test p-value			0.01	0.07	0.01

Notes: CAAR in percent (Short: [–1,+1], Medium: [–1,+5], Long: [–1,+20]). All CAAR values reported for high and low ESG groups. Differences significant at $p < 0.01$ (short and long windows) and $p = 0.07$ (medium window).

In the short window $[-1, +1]$, large-cap firms have an average CAR of -6.74%, whereas small-cap firms have -17.93%. This is a difference of 11.19 percentage points, which is highly significant ($p < 0.001$). Over the medium window $[-1, +5]$, large-cap firms experience an average CAR of -8.58% compared to -29.62% for small-cap firms, a gap of 21.04 pp ($p < 0.001$). By the long window $[-1, +20]$, the cumulative impact for large caps is -11.07% versus -26.97% for small caps, a difference of 15.90 pp ($p < 0.001$). These results indicate that larger firms indeed exhibit significantly milder stock price reactions to ESG incidents than smaller firms. Thus, the evidence supports **H3b**: firm size has a moderating effect, with large companies suffering substantially smaller abnormal losses relative to small companies when controversies occur.

Table 4.7. Firm size as a moderator (median split)

Firm Size Group	N	Short	Medium	Long	p-value
Large-cap	56	-6.74	-8.58	-11.07	
Small-cap	58	-17.93	-29.62	-26.97	
Difference (pp)		11.19	21.04	15.90	<0.001

Notes: CAAR in percent (Short: [–1,+1], Medium: [–1,+5], Long: [–1,+20]). All differences significant at $p < 0.001$ across all windows (t-tests).

4.4.3. Financial Health as a Moderator (H3c)

Finally, H3*c* considers aspects of financial health as potential moderators of the market reaction, specifically leverage (capital structure) and profitability. I examine each in turn by partitioning the sample at the median of the respective measure and comparing the CARs for high vs. low groups.

Leverage: No Moderating Effect

Leverage does not appear to significantly moderate the market reaction to ESG controversies. As shown in Table 4.8, after excluding a few firms with missing or zero leverage data, the sample for this test comprises 105 events (52 high-leverage firms and 53 low-leverage firms). High-leverage firms (above-median leverage) have a mean three-day CAR of -9.61%, compared to -15.23% for low-leverage firms. The difference of 5.62 percentage points is not statistically

significant ($p = 0.15$). This lack of a differential effect persists over longer horizons: in the medium window $[-1, +5]$, the high- vs. low-leverage CAR difference is 5.10 pp ($p = 0.38$), and in the long window $[-1, +20]$ it is 2.89 pp ($p = 0.60$). Across all event windows, leverage does not significantly distinguish the stock price reactions to controversies.

Profitability: Some Effect Observed

In contrast, profitability shows evidence of a moderating effect. Table 4.9 reports the results for firms split by profitability (ROA). High-ROA firms (above median ROA) experience a mean $[-1, +1]$ CAR of -8.72%, compared to -16.02% for low-ROA firms. This 7.30 pp difference is statistically significant ($p = 0.04$). For the medium window $[-1, +5]$, high-ROA firms have an average CAR of -14.62% vs. -23.79% for low-ROA, a gap of 9.17 pp that is marginally insignificant ($p = 0.09$). By the long window $[-1, +20]$, the difference between high- and low-profitability firms' CARs shrinks to 5.77 pp ($p = 0.29$) and is not significant.

Summary of H3c Results

In summary, among financial health variables, profitability provides some cushion against the negative impact of ESG controversies (especially in the immediate window), whereas leverage does not show a significant moderating influence. Stated differently, there is partial support for **H3c**: firms with stronger profitability metrics experience smaller short-term price declines following a controversy, but highly leveraged firms fare no better than low-leverage firms in these events.

Table 4.8. Leverage as a moderator (median split)

Leverage Group	N	Short	Medium	Long
High (above median)	52	-9.61	-16.88	-17.37
Low (below median)	53	-15.23	-21.97	-20.26
Difference (pp)		5.62	5.10	2.89
t-test p-value		0.15	0.38	0.60

Notes: CAAR in percent (Short: [–1,+1], Medium: [–1,+5], Long: [–1,+20]). The effective sample comprises 105 events after excluding observations with missing or zero leverage values. No differences are statistically significant across all windows ($p > 0.10$).

Table 4.9. Profitability as a moderator (median split)

ROA Group	N	Short	Medium	Long	p-value
High (above median)	56	-8.72	-14.62	-16.22	
Low (below median)	58	-16.02	-23.79	-21.99	
Difference (pp)		7.30	9.17	5.77	0.04*

Notes: CAAR in percent (Short: [–1,+1], Medium: [–1,+5], Long: [–1,+20]). *Short window significant at $p = 0.04$; medium window marginally significant at $p = 0.09$.

4.5. Robustness Checks

I conducted several robustness analyses to verify that the main findings are not sensitive to specific assumptions or sample quirks. The tests include handling of outliers and examining the stability of results over time.

4.5.1. Outliers and Winsorization

One concern is whether the negative CAAR results are driven by a few extreme events. To address this, I re-calculated the cumulative abnormal returns under different outlier treatments. Table 4.10 (Panel A) presents the results for the short window [-1,+1]. Winsorizing the individual CARs at the 1% level (i.e., limiting the most extreme values) yields a CAAR of -12.17% ($p < 0.001$). Using a 5% winsorization yields a CAAR of -10.94% ($p < 0.001$). Even with a more drastic approach of excluding the top and bottom 5% of CAR observations, the CAAR is -9.59% ($p < 0.001$). For the medium window $[-1, +5]$, the CAAR ranges from -18.85% (1% winsorized) to -15.25% (5% of observations excluded), and for the long window $[-1, +20]$ from about -18.99% to -16.32%. All of these estimates remain highly significant ($p < 0.001$). Therefore, a large negative market reaction is robust to outlier mitigation. This means that the effect is not driven by a small number of extreme cases.

4.5.2. Temporal Stability: Pre-2020 vs. 2020–2024

I also examined whether the market reactions have changed over time, by comparing an earlier subperiod (2015–2019) to the later part of the sample (2020–2024). For the short window $[-1, +1]$, pre-2020 events have an average CAAR of -11.80% ($p < 0.001$), while events from 2020 onward have an average CAAR of -13.11% ($p < 0.001$). This difference (about 1.3 percentage points) is not statistically significant ($p = 0.56$). A similar pattern holds for longer windows: for $[-1, +5]$, the mean CAR is -17.92% in the pre-2020 subsample vs. -20.75% in 2020–2024 (difference not significant at conventional levels), and for $[-1, +20]$ the figures are -16.99% vs. -21.49%. None of the differences between sub-periods are statistically significant. In short, there is no evidence of a structural break or time trend in the market's response to ESG controversies, and the negative reaction appears stable across the pre- and post-2020 samples.

Table 4.10. Robustness checks (CAR-based)

Check	Short [–1,+1]	Medium [–1,+5]	Long [–1,+20]
Baseline	**-12.43**	**-19.29**	**-19.16**
Winsorized (1%)	-12.17	-18.85	-18.99
Winsorized (5%)	-10.94	-17.60	-18.68
Exclude tails (5%)	-9.59	-15.25	-16.32
Pre-2020	-11.80	-17.92	-16.99
2020+	-13.11	-20.75	-21.49

Notes: CAAR in percent. All results significant at $p < 0.001$.

4.6. Summary of Findings and Hypothesis Support

This empirical results yield three primary findings, corresponding to the hypotheses:

First (H1): ESG controversies are associated with substantial shareholder value loss. Across the full sample of events, the CAARs are approximately -12.4% for the short window, -19.3% for the medium window, and -19.2% for the long window, and all are statistically significant at the $p < 0.001$ level.

Second (H2): The magnitude of the market impact differs by controversy type. In particular, governance-related incidents cause markedly larger stock price declines than environmental or social incidents (on the order of 2–3 times more negative CARs on average over a three-day window).

Third (H3): Firm-specific characteristics moderate the stock price impact of ESG controversies. A strong prior ESG reputation provides a significant buffering effect (high-ESG firms suffer only about 30% of the share price drop that low-ESG firms do), larger firm size is associated with much smaller CAR losses (large-cap firms experience only a fraction of the decline seen in small-cap firms), and higher profitability similarly cushions the impact. In contrast, leverage does not exhibit a significant moderating effect on the market reaction.

Collectively, these findings confirm that ESG controversy events have a materially negative impact on firm value, and the severity of this impact varies systematically depending on the nature of the event and the characteristics of the firm.

CHAPTER 5

Discussion

5.1. Main Findings

This thesis examines how global equity markets react when an ESG controversy becomes public. The results show a large, negative, and statistically robust price response, supporting H1. CARs are strongly negative in the immediate window and remain negative when the horizon is extended to several trading weeks. The reaction differs by controversy type, supporting H2. Governance incidents trigger substantially larger abnormal losses than environmental or social incidents, and this difference is visible in cross-sectional comparisons. Finally, firm characteristics shape the magnitude of the loss, supporting H3. In bivariate subgroup comparisons, firms with stronger pre-event ESG reputations, larger size, and higher profitability exhibit smaller abnormal losses. To address the concern that the ESG "buffer" merely proxies for fundamentals (e.g., size or profitability), the discussion emphasizes multivariate evidence: in cross-sectional regressions that jointly include ESG reputation and core firm controls, ESG remains positively associated with less negative CARs in the short and long windows (Appendix Table A.3). Collinearity is unlikely to drive this pattern given low VIFs (Appendix Table A.4). Leverage does not consistently increase the short-horizon reaction. Table 5.1 summarizes the main patterns.

Table 5.1. Summary of Key Empirical Results (N=114 Events)

Effect	**[-1,+1]**	**[-1,+5]**	**[-1,+20]**	**p-value**
H1: Overall CAR	-12.4%	-19.3%	-19.2%	< 0.001
H2: By Pillar				
Governance (G)	-20.4%	-29.1%	-27.4%	
Social (S)	-8.0%	-15.6%	-17.7%	Kruskal-Wallis
Environmental (E)	-7.9%	-10.6%	-9.1%	$p = 0.037$
H3a: ESG Reputation				
High ESG		-6.5%		$p = 0.01$
Low ESG		-22.9%		
H3b: Firm Size				
Large-cap		-6.7%		$p < 0.001$
Small-cap		-17.9%		
H3c: Profitability				
High ROA		-8.7%		$p = 0.04$
Low ROA		-16.0%		
H3d: Leverage		No effect		$p = 0.38$

Note: CARs from event study. p-values from t-tests (H1), Kruskal–Wallis (H2), subgroup t-tests (H3).

5.2. Comparison with Literature

The baseline finding is a negative abnormal return around ESG controversies. This finding aligns with a growing body of event study evidence. This evidence shows that capital markets penalize corporate misconduct and sustainability failures. Krüger (2015) shows that negative corporate responsibility events reduce shareholder wealth. Capelle-Blancard and Petit (2019) shows that ESG related news moves prices in economically meaningful ways. Using a broad international sample of controversies, De Winne and Petkeviciute (2022) reports negative abnormal performance after ESG controversies. More specific evidence in Aouadi and Marsat (2018) and Li et al. (2019) points in the same direction. Controversies are linked to losses in firm value and weaker market performance. Therefore, the negative CARs in this thesis are not unusual. They fit the main pattern that ESG failures are value relevant information.

Where my results differ is the size and persistence of the reaction. Several studies find negative announcement effects that are meaningful but smaller than the large CARs found here. One likely reason is the sample. Large database studies often include events with very different severity and media attention. When many events contain little new information, the average price response is smaller. In contrast, this thesis focuses on high visibility controversy events from 2015 to 2024. This design focuses on events where investors are more likely to update beliefs about future costs, governance quality, or the firm risk profile. In simple terms, a large average CAR is consistent with a sample that focuses on material controversies rather than minor allegations.

The ranking by pillar is a more distinctive result. Prior work often treats ESG news as one group. Other studies focus only on environmental or social issues. By comparing environmental, social, and governance events, I show a clear ranking. Governance incidents lead to the largest abnormal losses. This pattern is also visible in the cross sectional regression in Table 4.4. The governance indicator is linked to an additional CAR effect relative to the environmental baseline. Related evidence appears in Sun et al. (2024). That study links ESG scores, scandal probability, and event returns. It also shows that scandal type information is priced across firms. Overall, the results support the view that markets do not respond to ESG as one single concept. They respond to the specific content of the controversy.

This pillar ranking also connects to the materiality view of sustainability information. If governance incidents create the largest market response, one interpretation is that governance failures are seen as broadly material across industries. Environmental and social issues may depend more on the business model, regulation, and stakeholder exposure. This logic is consistent with Khan et al. (2016). That study shows that the market response is stronger for sustainability issues that are financially material. In controversy settings, governance problems are a direct signal about internal control, oversight, and ethical conduct. These issues are closely linked to cash flows and risk. It is therefore reasonable that investors treat them as high impact news. Environmental and social issues can also be highly material. However, investors may need more time or more context to translate them into cash flows and risk. This can reduce the average immediate effect in a mixed sample.

The moderation results link the thesis to work on reputational capital and downside protection. The idea that prior good conduct works like insurance against negative shocks appears in Godfrey (2005). It is also formalized as reputation insurance in Minor and Morgan (2011). In the bivariate moderator analyses, firms with higher pre-event ESG scores experience smaller abnormal losses than low-ESG firms, which is consistent with an "insurance-like" interpretation. Lins et al. (2017) provides related evidence in another setting, showing that CSR-related social capital improved firm performance during the financial crisis. Taken together, these arguments support the mechanism that trust and stakeholder relationships can reduce the equity value at risk when negative shocks arrive.

At the same time, interpreting a subgroup-based "ESG buffer" as causal is not straightforward. High-ESG firms are systematically different along dimensions that also predict event-time returns: they tend to be larger, more profitable, and unevenly distributed across controversy types and sectors. Median splits and bivariate subgroup tests therefore cannot fully isolate an ESG reputation effect from correlated fundamentals, raising a natural omitted-variable concern (e.g., whether the buffer is primarily a size/profitability proxy).

To address this identification risk, the thesis complements subgroup evidence with multivariate cross-sectional regressions that treat ESG reputation as a continuous predictor and jointly include a parsimonious set of firm fundamentals and event characteristics (log market capitalization, ROA, leverage, and pillar dummies; Appendix Table A.3). The controlled estimates continue to show a positive association between ESG reputation and CARs (i.e., less negative abnormal losses) in the short and long windows: a 10-point higher ESG score is associated with approximately 2.3 percentage points less negative CAR in $[-1, +1]$ and about 2.5 percentage points in $[-1, +20]$. The medium-window coefficient is smaller and not statistically distinguishable from zero, suggesting that any buffering is concentrated in the immediate market repricing and the longer adjustment rather than uniformly across horizons. Importantly, governance events remain associated with markedly more negative CARs even after controls, reinforcing the core heterogeneity result.

The regression design is intentionally parsimonious given the event-level sample size and the desire to avoid overfitting with many fixed effects; accordingly, the estimates should be interpreted as conditional associations rather than definitive causal effects. However, multicollinearity is unlikely to be driving the ESG coefficient: variance inflation factors are low (all below 2; Appendix Table A.4), indicating that ESG reputation is not mechanically redundant with the included controls.

Finally, ESG reputation is not a simple measure. Recent work shows that ESG scores differ substantially across rating providers. Berg et al. (2022) documents large differences in ESG ratings and shows that ratings reflect different choices and assumptions. This matters for moderation results because a high ESG score may proxy for a broader information environment (e.g., disclosure intensity or governance systems) and provider-specific weighting schemes. The key point in this thesis is therefore empirical and conditional: using the selected ESG measure, and

controlling for core fundamentals, markets react less negatively for firms with stronger pre-event ESG reputations.

The size and profitability moderators fit a broader view that firm resources shape how markets price negative shocks. Cui and Docherty (2020) finds that stock return reactions to ESG related information differ across firms. This is consistent with differences in how investors update beliefs. More broadly, evidence that corporate social performance is linked to risk and resilience appears in Jo and Na (2012) and Oikonomou et al. (2012). In my setting, size and profitability likely capture capacity. This includes the ability to fund remediation, manage legal and regulatory exposure, keep key stakeholders, and communicate well under scrutiny. These factors can reduce expected tail costs and reduce the time of uncertainty that investors price.

Finally, the absence of a leverage moderator is informative. It does not support a simple balance sheet amplification story in short event windows. Highly leveraged firms do not show systematically larger abnormal losses than low leverage firms. One interpretation is that the short run equity reaction is driven by expected stakeholder and governance costs. These may dominate any immediate financing channel. Another interpretation is that leverage effects appear first in credit markets. Equity markets may mainly price the expected value impact of the controversy itself. Either way, the null result is useful. It narrows the set of mechanisms that fit the observed short horizon equity response.

5.3. Theoretical Implications

The results provide evidence on how quickly and through which channels ESG-related information is incorporated into equity prices. The sharp abnormal return response around controversy dates is consistent with semi-strong market efficiency: when new information becomes public, prices adjust rapidly (Fama, 1970). The persistence of negative CARs over multi-week windows suggests that the market views controversies as news with lasting valuation relevance rather than transitory noise. In event-study terms, the market response looks like an information update that affects expected value and risk, not only temporary sentiment.

A useful way to interpret the CARs is through the two standard valuation levers: expected cash flows and the discount rate. A controversy can reduce expected cash flows via fines, compliance costs, disrupted production, and lost business opportunities. It can also change the discount rate by increasing perceived risk, especially downside risk and uncertainty around future liabilities. The valuation perspective in Giese et al. (2019) emphasizes both channels: ESG can affect expected cash flows through stakeholder and regulatory outcomes and can affect discount rates through risk premia. The results are consistent with this framework because the market response is large and remains negative through the 20-trading-day window, suggesting that investors do not treat controversies as transitory noise.

The governance premium in the results aligns naturally with agency theory. Governance controversies deliver direct negative information about monitoring and the alignment between managers and shareholders (Jensen and Meckling, 1976). When investors learn about fraud, corruption, or severe compliance failures, they update beliefs about future agency costs and the

likelihood of repeated misconduct. This update can be large because governance failures do not only imply one-off penalties; they also signal weaknesses in internal controls and ethical standards. These weaknesses raise uncertainty about future decisions and disclosures. The strong market penalty is therefore consistent with a story in which governance controversies increase both expected costs and perceived risk.

A legitimacy perspective helps explain why controversies in any pillar trigger negative reactions. A controversy is a public signal that the firm violated social expectations. This can threaten its legitimacy and its ability to maintain stakeholder support. Suchman (1995) describes legitimacy as a generalized perception that organizational actions are appropriate within a socially constructed system of norms. From this viewpoint, the negative abnormal returns reflect the market's expectation that a legitimacy shock reduces future value, whether through customer responses, regulatory pressure, labor-market effects, or contracting frictions. The key empirical addition from this thesis is that legitimacy shocks are not priced uniformly across ESG dimensions; governance-related legitimacy shocks appear especially costly.

The moderation results point to signaling and moral capital. In signaling models, investors interpret new information in the context of prior signals and reputations (Spence, 1973; Connelly et al., 2011). A strong ESG reputation can reduce the "surprise" of a controversy by shifting beliefs about the firm's underlying type. If investors view a high-ESG firm as less likely to be systematically irresponsible, then a single controversy provides weaker evidence of persistent problems. Conversely, when a firm has a weak ESG record, a controversy may confirm prior beliefs and trigger a larger belief revision. This belief-updating perspective fits the observed pattern: the market penalty is much larger for low-ESG firms.

This signaling logic also connects to the "insurance" view of CSR. If prior ESG performance builds moral capital, stakeholders may respond less harshly when bad news arrives (Godfrey, 2005; Minor and Morgan, 2011). The mechanism does not require stakeholder forgiveness in a normative sense; it can operate through expectations. Investors may expect that customers, employees, and regulators give high-reputation firms more time to remediate, reducing the probability of extreme outcomes. The evidence in Lins et al. (2017) supports the broader idea that trust and stakeholder relationships improve resilience under adverse conditions. In this thesis, the ESG reputation buffer is consistent with a market belief that high-reputation firms face lower expected tail costs from controversies.

Firm size and profitability can be interpreted through an organizational resilience lens. These characteristics proxy for slack resources and managerial capacity. A large firm often has diversified operations, specialized compliance functions, and access to capital. A profitable firm can finance remediation and absorb direct costs without threatening operations. These features can lower the expected tail costs of a controversy and can speed resolution of uncertainty, which reduces the discount-rate impact. This interpretation is consistent with the broader risk channel emphasized in the CSR-risk literature (Jo and Na, 2012; Oikonomou et al., 2012): firms differ in their capacity to absorb shocks, and markets price that difference.

The leverage null result is theoretically useful because it implies that short-horizon equity pricing of controversies is not strongly mediated by capital structure. Leverage may still matter for long-run recovery, but within a few weeks the equity market response appears to be driven more by governance quality and stakeholder costs. This highlights a broader point: different aspects of a controversy may be priced in different markets and at different horizons. Equity prices react quickly to changes in expected value; financing constraints and covenant effects may become more salient later.

Finally, measurement is part of theory. ESG ratings divergence implies that the same firm can be classified differently depending on the data provider. Berg et al. (2022) shows that divergence is substantial and can create "aggregate confusion." In this setting, an ESG score is best viewed as a noisy signal. The moderating role of ESG reputation in my results therefore supports the idea that investors treat ESG assessments as informative, but imperfect, inputs into belief formation. That nuance matters for interpretation: the buffer effect is evidence that markets respond to ESG reputation proxies, not proof that any single rating measures "true" ESG quality without error.

5.4. Practical Implications

The findings matter for investors, managers, and regulators because they quantify the market value at risk in ESG controversies and clarify which characteristics shape that risk.

Implications for investors. ESG controversies behave like large negative shocks rather than small news items. The average abnormal loss is economically meaningful even for diversified portfolios, and it becomes especially relevant for investors with concentrated exposures or sector tilts. The persistence of negative CARs over several weeks suggests that markets interpret controversies as events with real valuation consequences, not just temporary sentiment. For valuation practice, this means that controversy risk is not only a "reputation" concern. It is a cash-flow and discount-rate risk that investors should model explicitly, consistent with the valuation channels discussed in Giese et al. (2019).

The pillar results sharpen risk monitoring. Governance controversies are punished most severely, so investors may benefit from giving governance indicators special weight, including internal control quality, enforcement history, and board oversight. This is a practical complement to the materiality approach in Khan et al. (2016): investors can focus attention on issues that are most likely to affect firm value across industries. The results do not imply that environmental and social risks are immaterial. Instead, they indicate that, in an average short-horizon reaction, governance failures carry the strongest immediate signal for valuation and risk.

The reputation-buffer result also has portfolio implications. Firms with stronger ESG reputations experience smaller abnormal losses, which suggests that ESG quality can provide partial downside protection. This is consistent with the risk-management view of CSR (Godfrey, 2005) and evidence linking social performance to risk (Oikonomou et al., 2012). At the same time, investors should avoid mechanical reliance on a single ESG rating. Given the rating divergence in

Berg et al. (2022), a reasonable practical approach is triangulation: combine ESG ratings with governance indicators, disclosure quality assessments, and controversy histories.

Implications for managers and boards. For corporate decision-makers, the most direct implication is speed and scale: controversies can erase a meaningful share of market value quickly. Governance failures are especially costly, which implies that investments in compliance, audit quality, and ethical oversight are central to value protection. Boards can read the governance premium as evidence that investors treat governance integrity as a prerequisite for trusting management, disclosures, and strategy. This strengthens the business case for investing in internal controls and for actively monitoring governance risk as part of enterprise risk management.

The moderation results suggest a proactive strategy. Firms that enter a controversy with strong reputations and strong fundamentals are punished less. This points to the value of building resilience before a shock occurs. The strongest interpretation is not that a firm can "buy" forgiveness through branding, but that credible, consistent ESG performance and transparent governance systems build trust and reduce uncertainty when bad news arrives. This fits the reputation insurance logic in Minor and Morgan (2011). Managers who treat ESG as part of risk management, not only as reporting, may reduce the equity value at risk during a controversy.

Implications for regulators and standard setters. The evidence that markets impose large penalties after ESG controversies supports the view that controversies are financially material information. This strengthens the case for timely, comparable disclosure and for enforcement that improves the credibility of ESG reporting. Standardization can reduce information asymmetry, especially for investors who cannot independently verify ESG quality. It can also reduce the reliance on fragmented private ratings, which is important given documented divergence across rating providers (Berg et al., 2022). From a governance perspective, the results also suggest that enforcement that improves governance integrity can have first-order effects on investor confidence.

5.5. Limitations

Several limitations qualify the interpretation of the results and point to where the evidence is strongest and where it should be read more cautiously.

Sample scope and external validity. The sample contains 114 controversy events from listed firms over 2015–2024 and is skewed toward large firms. This design is appropriate for high-visibility controversies but limits generalizability. Smaller firms may show different patterns because they have thinner trading, less analyst coverage, and different investor bases. The sample also focuses on material controversies, which likely increases the estimated average effect relative to studies that include minor allegations. The results therefore best describe market reactions to major controversies in a large-firm setting.

Event definition and confounding news. Event studies assume that the event date captures the arrival of new information and that other news does not systematically contaminate the window. In practice, controversies can unfold over several days and can coincide with other announcements. This creates a risk that part of the CAR reflects correlated information rather than

the controversy itself. Short windows reduce this risk but do not eliminate it. This limitation is common in event-study settings and is a key reason why careful event-screening and robustness analysis are central to credible inference (MacKinlay, 1997; Brown and Warner, 1985).

Short-horizon focus: The analysis uses windows up to 20 trading days. This improves attribution but does not speak to long-run recovery or long-run operating effects. Longer-horizon designs face well-known econometric issues and lower power (Kothari and Warner, 2006). The persistence of negative CARs through day +20 suggests slow resolution of uncertainty, but it is not evidence of permanent value loss.

Inference and dependence: Statistical inference in event studies can be affected by cross-sectional dependence when multiple events cluster in calendar time or share common exposure (e.g., industry shocks), which may inflate conventional t-statistics (Kolari and Pynnonen, 2010). To address this concern, I complement standard one-sample t-tests with calendar-date clustered inference (Appendix A); the resulting test statistics are comparable to the baseline results. In addition, because event-level CARs may deviate from normality, I report non-parametric robustness tests alongside parametric inference (Boehmer et al., 1991; Corrado and Zivney, 1992). While these procedures improve robustness, no single test fully eliminates all dependence and distributional concerns.

Measurement of pillars and ESG reputation: Each event is classified into a primary ESG pillar even though some events plausibly span multiple dimensions. This classification enables comparison but can blur mixed events. The ESG reputation measure relies on ESG scores that diverge across providers (Berg et al., 2022). The high- vs. low-ESG split captures meaningful heterogeneity, but it remains a coarse proxy for reputational capital and underlying systems.

The final event set includes manual validation of event dates and salience to ensure that Day 0 reflects the initial significant public disclosure. Although this step improves the precision of the event dates, it introduces researcher discretion. Therefore, the findings are best interpreted as evidence of major ESG controversy disclosures rather than minor ESG-related incidents.

5.6. Future Research

The results open several directions for future research.

Link short-window reactions to longer-run outcomes: Future work can connect short-window CARs to medium- and long-run outcomes, such as changes in profitability, investment, litigation costs, or risk measures. This would help separate fundamental repricing from temporary overreaction and would clarify which channels drive the initial equity response.

Study credit-market responses and the leverage null: The absence of leverage moderation in short-window equity returns is a natural puzzle. One approach is to test whether leverage matters more in credit markets than in equities. If controversies increase perceived default risk or covenant pressure, bond spreads or ratings may respond even when equity CARs do not differ by leverage.

Refine controversy classification and incorporate materiality: The governance premium motivates a more granular taxonomy of governance events and a closer link to materiality. Future work can build on Khan et al. (2016) by testing whether market reactions are larger when a controversy maps to financially material topics for the firm's industry, and whether the governance premium is driven by specific subtypes of failures.

Improve measurement of ESG reputation and the information environment: Given ESG ratings divergence (Berg et al., 2022), future research can test whether the reputation buffer is stronger when ESG assessments are more consistent across providers, when disclosure quality is higher, or when controversies are more surprising given past signals. Modeling the joint relation between ESG scores, scandal probability, and event returns, as in Sun et al. (2024), is a promising way to formalize this information-environment channel.

CHAPTER 6

Conclusion

Equity markets penalize ESG controversy disclosures sharply. In a sample of 114 high-visibility controversy events (2015–2024), the average CAAR equals -12.43% in the short window $[-1, +1]$, deepens to -19.29% over $[-1, +5]$, and remains at -19.16% through $[-1, +20]$ (Table 4.3). Governance controversies drive the deepest value losses: -20.41% in $[-1, +1]$ versus roughly -8% for environmental and social events (Table 4.5). Firm characteristics reshape the penalty. High-ESG firms lose -6.53% while low-ESG firms lose -22.88% in $[-1, +1]$ (Table 4.6); large-cap firms lose -6.74% versus -17.93% for small caps (Table 4.7); and high-profitability firms lose -8.72% versus -16.02% for low-profitability firms (Table 4.9). Leverage does not moderate the short-run reaction (Table 4.8). The answer to the research question is direct: ESG controversies erase value quickly, governance failures erase the most, and credibility plus capacity determine how much value remains at risk.

6.1. Comprehensive Synthesis of Findings

The results group into four clusters. Together, they form one coherent story about how markets process negative ESG information.

Cluster 1: the market reprices quickly within the short horizon: Abnormal losses appear immediately and remain economically large across all windows (Table 4.3). The incremental change from $[-1, +1]$ to $[-1, +5]$ accounts for most of the cumulative decline, while the change from $[-1, +5]$ to $[-1, +20]$ is small. This timing is consistent with rapid incorporation of controversy-related information within the first post-event week and with no evidence of a price reversal within the 20-trading-day horizon. Because the $[-1, +20]$ window can include follow-up disclosures and other contemporaneous firm news, the long-window CAAR is interpreted as sustained negative performance over the examined horizon rather than as proof of a permanent effect.

Cluster 2: governance dominates the information content: Environmental and social controversies trigger significant losses, but governance incidents trigger the largest losses in every window (Table 4.5). The regression evidence reinforces this ranking: in the medium window, governance events load as an additional -18.48 percentage points relative to the environmental baseline (Table 4.4). This pattern clarifies how investors differentiate across ESG dimensions. Governance failures function as system-level signals about oversight quality and reporting credibility. Investors price them as indicators of persistent control problems rather than isolated operational mishaps.

Cluster 3: resilience is priced through credibility and resources: The market penalty varies sharply with pre-event credibility and with operational capacity. High ESG reputation buffers the reaction (Table 4.6). The size effect is equally strong and persists over longer windows (Table 4.7). Profitability cushions the immediate loss (Table 4.9). These patterns identify an economically relevant interaction: when bad news arrives, investors do not only price the incident; they also price the firm's ability to absorb direct costs, manage stakeholders, and resolve uncertainty. Resilience shows up in returns.

Cluster 4: robustness and the limits of balance-sheet amplification: The negative CAAR survives winsorization and tail trimming and remains stable across pre-2020 and 2020–2024 subsamples (Table 4.10). The effect is not driven by a few extreme events or by one unusual period. At the same time, leverage does not moderate the reaction (Table 4.8). This null result matters. It narrows the short-run mechanism: equity markets punish expected stakeholder and governance costs more than immediate balance-sheet fragility.

Together, these clusters solve the core problem. ESG controversies act as material valuation shocks. Markets differentiate across controversy content and across firm resilience, and that differentiation is systematic and economically large.

6.2. Theoretical Contributions

This thesis contributes to theory by connecting established mechanisms to precise empirical patterns.

First, the timing profile supports semi-strong market efficiency in an ESG context. In an efficient market, public value-relevant information should enter prices quickly (Fama, 1970; Fama et al., 1969). The evidence is consistent with rapid price adjustment: most of the cumulative movement occurs by the first post-event week, and CAAR remains similar through day $+20$ (Table 4.3). However, because longer windows may incorporate follow-up disclosures, the results cannot fully separate the initial announcement effect from subsequent information releases.

Second, the governance premium operationalizes agency and internal-control theories in the ESG controversy setting. Agency theory predicts that governance failures increase expected agency costs and uncertainty (Jensen and Meckling, 1976). The results show that governance incidents receive the largest and most persistent penalties (Table 4.5), and governance status explains variation in CAR even in a regression framework (Table 4.4). This evidence strengthens a specific interpretation of ESG materiality: governance failures convey high-information content about future downside risk because they undermine the credibility of monitoring and disclosure.

Third, the results resolve the reputation *buffer versus backlash* debate for this setting. Signaling theory implies that investors interpret new information in the context of prior reputation (Spence, 1973; Connelly et al., 2011). The data show a strong buffer: high-ESG firms suffer far smaller losses than low-ESG firms (Table 4.6). This supports the "insurance" view of CSR and reputational capital (Godfrey, 2005; Minor and Morgan, 2011; Lins et al., 2017). It also contradicts a greenwashing-driven backlash mechanism as the dominant force in this sample (Yu

et al., 2020; Sun et al., 2024). Importantly, the result holds even though ESG scores are imperfect signals and differ across providers (Berg et al., 2022). Investors still price the information embedded in reputation proxies.

6.3. Practical and Managerial Implications

The estimates translate into actions because they quantify downside risk and identify where it concentrates. The penalty is steep. Governance is the fault line.

Investors and risk owners: ESG controversies behave like event risk, not background noise. First, build an event-monitoring process that classifies incidents by pillar and severity and links them to pre-defined portfolio rules; use estimated CAR magnitudes as stress-test inputs rather than generic ESG labels (Tables 4.3 and 4.5). Second, reweight governance diagnostics in due diligence and engagement, because governance controversies carry the largest valuation losses. Third, price tail risk asymmetrically. Low-ESG and small-cap firms experience the sharpest abnormal losses (Tables 4.6 and 4.7); portfolio limits, liquidity planning, and required returns should reflect that distribution.

CEOs, boards, and control functions: A controversy becomes a valuation event on day zero. Prevention therefore has a measurable payoff. Firms should treat internal controls, audit quality, escalation procedures, and whistleblowing protections as value protection rather than overhead. The governance premium shows that investors interpret control failures as repeat-risk signals. Execution matters as much as prevention. When an incident occurs, management should reduce uncertainty with rapid disclosure, an independent investigation, and a quantified remediation plan with deadlines. Investors punish silence and vagueness because they prolong uncertainty about cash-flow impact and enforcement risk. Credibility also pays before a crisis. High-ESG firms lose less because they enter the event with accumulated trust and process quality (Table 4.6). Firms build that trust through consistent performance and verifiable targets, not through marketing narratives.

Regulators and standard setters: The observed price penalties demonstrate financial materiality. Policy should prioritize timeliness and comparability of incident disclosure to reduce information asymmetry and rumor-based price discovery. Enforcement should raise the governance floor, because governance incidents impose the largest valuation losses. Standard setters should also improve transparency in ESG data markets. Divergent ratings create "aggregate confusion" (Berg et al., 2022); clearer methodologies, auditability, and disclosure of key assumptions strengthen how markets interpret ESG reputation signals.

6.4. Limitations and Future Research

This thesis draws strong inference where the design is strongest: short-horizon valuation effects around precisely dated, high-visibility controversy disclosures (Brown and Warner, 1985; MacKinlay, 1997). The same design sets boundaries. The sample is limited to 114 events from 2015–2024 (Table 4.1), and it is skewed toward large firms, so the estimates describe material incidents in a large-firm setting. Event timing also remains a constraint. Controversies can unfold over several days and coincide with other corporate news, so short windows reduce confounding but do not eliminate it (Kothari and Warner, 2006). Finally, pillar assignment uses a primary category even when incidents span dimensions, and ESG reputation relies on provider scores that contain measurement noise (Berg et al., 2022).

Two next-step studies follow directly. First, link short-window CARs to *real* outcomes using a matched-sample panel design: estimate difference-in-differences effects on profitability, investment, and litigation costs over one to three years, and decompose price responses into cash-flow versus discount-rate revisions using analyst forecast updates and implied costs of capital. Second, test the leverage null in adjacent markets by running parallel event studies on bond spreads or CDS premia; credit markets should price balance-sheet fragility more directly than equities, and that comparison would isolate the financing channel that is muted in the short-run equity response.

ESG controversies are priced shocks, not soft narratives. This thesis quantifies the shock and shows where it concentrates: governance failures, weak credibility, and limited capacity. It also shows what protects value: disciplined governance systems and earned ESG reputation. Markets do not wait for annual reports. They reprice on impact. Firms and investors who treat ESG control as core risk management keep more shareholder value when the next controversy breaks.

References

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.

Berg, F., Kölbel, J. F., and Rigobon, R. (2022). Aggregate confusion: The divergence of esg ratings. *Review of Finance*, 26(6):1315–1344.

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

Brighi, P., Lucarelli, C., and Marinelli, N. (2023). Esg controversies and bank risk-taking. *Business Strategy and the Environment*, 32(1):274–288.

Brown, S. J. and Warner, J. B. (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics*, 14(3):205–258.

Capelle-Blancard, G. and Petit, A. (2019). Every little helps? esg news and stock market reaction. *Journal of Business Ethics*, 157(2):543–565.

Connelly, B. L., Certo, S. T., Ireland, R. D., and Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, 37(1):39–67.

Corrado, C. J. and Zivney, T. L. (1992). The specification and power of the sign test in event study hypothesis tests using daily abnormal returns. *Journal of Financial and Quantitative Analysis*, 27(3):465–478.

Cui, L. and Docherty, P. (2020). Esg analysis and stock return: An event study approach. In *World Finance Conference*, Budapest, Hungary. Conference proceeding.

De Winne, K. and Petkeviciute, N. (2022). Does esg controversy hurt stock performance? an international event study. *Sustainability*, 14(10):5953.

Elamer, A. A. and Boulhaga, M. (2024). Esg controversies and corporate performance: The moderating effect of governance mechanisms and esg practices. *Corporate Social Responsibility and Environmental Management*, 31(4):3312–3327.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417.

Fama, E. F., Fisher, L., Jensen, M. C., and Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, 10(1):1–21.

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.

Frank, M. Z. and Goyal, V. K. (2009). Capital structure decisions: Which factors are reliably important? *Financial Management*, 38(1):1–37.

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

Giese, G., Lee, L.-E., Melas, D., Nagy, Z., and Nishikawa, L. (2019). Foundations of esg investing: How esg affects equity valuation, risk, and performance. *Journal of Portfolio Management*, 45(5):69–83.

Global Sustainable Investment Alliance (2023). Global sustainable investment review 2022. Report.

Godfrey, P. C. (2005). The relationship between corporate philanthropy and shareholder wealth: A risk management perspective. *Academy of Management Review*, 30(4):777–798.

Huang, K.-J., Bui, D. G., Hsu, Y.-T., and Lin, C.-Y. (2024a). The esg washing in banks: Evidence from the syndicated loan market. *Journal of International Money and Finance*, 142:102008.

Huang, Q., Li, Y., and Lin, M. (2024b). Esg washing in banking: Evidence from funding patterns. *Journal of Banking & Finance*. Forthcoming.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305–360.

Jo, H. and Na, H. (2012). Does csr reduce firm risk? evidence from controversial industry sectors. *Journal of Business Ethics*, 110(4):441–456.

Karpoff, J. M., Lee, D. S., and Martin, G. S. (2008). The cost to firms of cooking the books. *Journal of Financial and Quantitative Analysis*, 43(3):581–611.

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

Kolari, J. W. and Pynnonen, S. (2010). Event study testing with cross-sectional and time-series dependence. *Journal of Financial and Quantitative Analysis*, 45(5):1319–1343.

Kothari, S. P. and Warner, J. B. (2006). Econometrics of event studies. In *Handbook of Corporate Finance: Empirical Corporate Finance*, volume 1, chapter 1, pages 3–36. North-Holland/Elsevier.

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

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621.

Li, J., Haider, Z. A., Jin, X., and Yuan, W. (2019). Corporate controversy, social responsibility, and market performance. *Journal of International Financial Markets, Institutions and Money*, 60:101118.

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. *Journal of Finance*, 72(4):1785–1824.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, 35(1):13–39.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60.

Minor, D. and Morgan, J. (2011). Csr as reputation insurance: Primum non nocere. *California Management Review*, 53(3):40–59.

Mu, X. and Shi, Y. (2023). An event study of the impact of negative esg news on stock returns. In *Proceedings of the International Conference on Business and Information Systems (ICBIS)*, volume 14 of *Advances in Human-Centered and Social Computing*, pages 1384–1398. Springer.

Nirino, N., Santoro, G., Miglietta, N., and Quaglia, R. (2021). Corporate controversies and company's financial performance: Exploring the moderating role of esg practices. *Technological Forecasting and Social Change*, 162:120341.

Oikonomou, I., Brooks, C., and Pavelin, S. (2012). The impact of corporate social performance on financial risk and utility: A longitudinal analysis. *Financial Management*, 41(2):483–515.

Rock, E. B. (2022). Understanding the role of esg and stakeholder governance within the framework of fiduciary duty. *Harvard Law School Forum on Corporate Governance*.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374.

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3):571–610.

Sun, W., Luo, Y., Yiu, S.-M., Yu, L., and Ding, W. (2024). Esg scores, scandal probability, and event returns. *Financial Innovation*, 10:121.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, 2 edition.

Yu, P.-Y., Luu, B. V., and Chen, C.-H. (2020). Greenwashing in environmental, social and governance disclosures. *Research in International Business and Finance*, 52:101192.

APPENDIX A

Appendix

Table A.1. Benchmark index mapping by primary listing market

Listing market	Suffix	Benchmark (RIC)	Benchmark name
Brazil (B3)	SA	.BVSP	Brazilian IBOVESPA
Canada (TSX)	TO	.GSPTSE	S&P/TSX Composite
China (Shenzhen)	SZ	.SSEC	Shanghai Composite (proxy)
Denmark (Nasdaq Copenhagen)	CO	.OMXC20	OMX Copenhagen 20
France (Euronext Paris)	PA	.FCHI	CAC 40
Germany (Hamburg)	H	.GDAXI	DAX
Germany (Xetra/Frankfurt)	DE	.GDAXI	DAX
Hong Kong (HKEX)	HK	.HSI	Hang Seng
India (NSE)	NS	.NSEI	NIFTY 50
Israel (TASE)	TA	.TA35	TA-35 (Tel Aviv 35)
Japan (TSE)	T	.TOPX	TOPIX
Malaysia (Bursa Malaysia)	KL	.KLSE	FTSE Bursa Malaysia KLCI
Netherlands (Euronext Amsterdam)	AS	.AEX	AEX
Norway (Oslo Børs)	OL	.OBX	Oslo OBX
Russia (MOEX)	MM	.IRTS	RTS Index
Singapore (SGX)	SI	.STI	Straits Times Index
South Africa (JSE)	J	.FTJ203	JSE All Share
South Korea (KRX)	KS	.KS11	KOSPI Composite
Spain (BME Madrid)	MC	.IBEX	IBEX 35
Sweden (Nasdaq Stockholm)	ST	.OMXS30	OMX Stockholm 30
Switzerland (SIX)	S	.SSMI	Swiss Market Index
Taiwan (TWSE)	TW	.TWII	TWSE Capitalization Weighted Index
United Kingdom (LSE)	L	.FTSE	FTSE 100
United States (Consolidated)	K	.SPX	S&P 500
United States (NASDAQ)	O	.SPX	S&P 500
United States (NASDAQ)	OQ	.SPX	S&P 500
United States (NYSE)	N	.SPX	S&P 500
United States (OTC)	PK	.SPX	S&P 500
Australia (ASX)	AX	.AORD	ASX All Ordinaries

Notes: Suffix denotes the Refinitiv RIC exchange extension. Benchmarks are Refinitiv index RICs; unmapped cases use .MIWD00000PUS. Shenzhen listings use .SSEC as a proxy when needed.

Table A.2. Adjusted inference for CAAR (calendar-date clustered)

Window	N	CAAR (%)	Standard t (p-value)	Clustered-by-date t (p-value)
Short [–1,+1]	114	–12.43	–6.79 (< 0.001)	–6.71 (< 0.001)
Medium [–1,+5]	114	–19.29	–7.08 (< 0.001)	–7.12 (< 0.001)
Long [–1,+20]	114	–19.16	–7.09 (< 0.001)	–7.50 (< 0.001)

Notes: Standard t-tests are one-sample tests of event CARs against zero. Clustered t-tests use calendar-date clustered standard errors (109 unique dates). Wilcoxon tests reject zero median CAR for all windows ($p < 0.001$).

Table A.3. Cross-sectional regressions with controls (ESG reputation and firm fundamentals)

	Short		Medium		Long	
Variable	Coef.	(SE)	Coef.	(SE)	Coef.	(SE)
ESG score (per 10 points)	0.0228**	(0.0097)	0.0147	(0.0139)	0.0250*	(0.0141)
Log market cap (t−1)	0.0073	(0.0078)	0.0163	(0.0113)	0.0108	(0.0121)
ROA (t−1)	0.4680*	(0.2656)	0.3158	(0.3307)	0.1813	(0.3545)
Leverage (t−1)	0.1225	(0.1024)	-0.0049	(0.0872)	-0.0476	(0.1050)
Pillar: Social	0.0050	(0.0260)	-0.0356	(0.0496)	-0.0389	(0.0447)
Pillar: Governance	-0.1186**	(0.0520)	-0.1472**	(0.0634)	-0.1493**	(0.0648)
Constant	-0.4731**	(0.2083)	-0.6090**	(0.2569)	-0.5061*	(0.2823)
R^2	0.206		0.120		0.134	
Adj. R^2	0.159		0.067		0.082	
N	108		108		108	

Notes: Dependent variable is event-level CAR in each window (Short: [–1,+1], Medium: [–1,+5], Long: [–1,+20]). ESG score is scaled per 10 points. E-pillar is the baseline category. Heteroskedasticity-robust standard errors (HC3). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A.4. Variance inflation factors (VIF) for ESG-control regressions

Variable	VIF
ESG score (per 10 points)	1.12
Log market cap (t−1)	1.19
ROA (t−1)	1.24
Leverage (t−1)	1.16
Pillar: Social	1.62
Pillar: Governance	1.60

Notes: VIFs are computed for the regressors used in Table A.3.